САНКТ-ПЕРЕРБУРГКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Сахаров Вадим Юрьевич

Лекции по теории вероятностей

Учебное пособие

Санкт-Петербург
2018

УДК 519.2
ББК 22.171

Рецензенты: к.ф.-м.н., доц. С.М.Ананьевский (С.-Петерб. гос. ун-т), к.ф.-м.н., доц. А.А.Кельзон (Гос. ун-т морского и речного флота им адм. С.О.Макарова).

Рекомендовано к опубликованию
учебно-методической комиссией
математико-механического факультета
С.-Петербургского государственного университета

Сахаров В.Ю.

Лекции по теории вероятностей: учеб. пособие. – СПб.: СПбГУ, 2018. – 87 с.

Учебное пособие написано на основе лекций, читавшихся автором на Биолого-почвенном Геологическом факультетах, а также отделений Почвоведения и Нефтегазового дела (ныне включённых в Институт наук о Земле) Санкт-Петербургского Государственного университета в течение двух десятков лет. В книге без использования традиционных для математики строгостей, излагаются основные принципы теории вероятностей. Некоторые из вопросов имеют оригинальную трактовку. В книге присутствуют некоторые параграфы, которых нет в других учебниках, описано актуальное в наше время компьютерное моделирование случайных величин.

Для студентов нематематических специальностей.

Предисловие

Во время преподавания математики студентам гуманитарных специальностей возникает следующая проблема: в отличие от студентов-математиков им не свойственно безоговорочно принимать базовые утверждения: определения и аксиомы. Действительно, математик, определяя некоторые понятия (он их называет *абстракциями*), не доказывает, что они совпадают с какими-то реальностями в жизни. Совершенно естественно, что гуманитарии часто задают бессмысленный, с точки зрения математика, вопрос: «Почему определение именно такое? Докажите определение!». Конечно доказать определение нельзя, но всё-таки возможно найти сходства математических понятий и окружающих нас вещей.

Изложенный ниже материал не претендует на строгость. Доказательства теорем (а иногда и их строгие формулировки) практически отсутствуют. Поэтому прочтение данного труда вряд ли будет полезно студентам, специализирующимся в области математики. А гуманитарии, наделённые «здравым смыслом», смогут найти здесь желанные комментарии и объяснения.

Теория случайных событий

События

Будем считать, что поводится некоторый эксперимент. Само проведение эксперимента будем называть *испытанием*. Результат этого эксперимента будем называть *событием*. Если событие происходит при каждом испытании, то его будем называть *достоверным* и обозначать греческой буквой Ω (омега). Если событие не наступает ни при каком испытании, то его будем называть *невозможным* и обозначать символом пустого множества: $\varnothing$. Если событие может произойти, а может и не произойти, то его будем называть *случайным* и обозначать большими латинскими буквами: A, B, C, … .

Заметим, что использование символики теории множеств здесь вполне оправдано. Кроме символа пустого множества в дальнейшем будут использоваться равенства и подмножества, знаки операций пересечения, объединения и разности множеств. Так, если события A и B всегда наступают одновременно, то будем записывать $A=B$. Если наступление события A влечёт за собой наступление события B, то это обозначим как $A\subset B$. Событие A в этом случае называется *подсобытием* (по аналогии с термином *подмножество*) события B. Событие, наступающее при одновременном наступлении событий A и B, будем обозначать $A\cap B$. Событие, которое наступает, в тот случае, если произошло <u>хотя бы</u> одно из событий A или B, будем обозначать $A\cup B$. Событие, которое наступает, если событие A произошло и при этом событие B не произошло, обозначим $A\backslash B$. Аналогом достоверного события Ω в теории множеств является *универсальное множество*. *Противоположным событием* для события A будем называть событие $\Omega\backslash A$, которое обозначим чёрточкой наверху: $\overline{A}=\Omega\backslash A$ (эта операция аналогична дополнению в теории множеств). Справедливы равенства $A\cap\overline{A}=\varnothing$, $A\cup\overline{A}=\Omega$.

События, которые не могут наступить одновременно, называются *несовместными*. Если события A и B несовместны, то $A\cap B=\varnothing$.

Иногда бывает удобно использовать понятие *элементарного* события. Элементарными называются события для некоторого эксперимента, не содержащие никаких подсобытий, кроме невозможного события и самого себя. Событие A будет элементарным для данного эксперимента, если для любого события B справедливо одно из утверждений: $A\subset B$ или $A\cap B=\varnothing$. То есть, либо событие A целиком входит в событие B, либо целиком не входит. Множество, состоящее из всех элементарных событий данного эксперимента, называется *вероятностным пространством*.

В качестве примера эксперимента приведём бросание игральной кости (кубика). При таком эксперименте можно фиксировать число очков на его верхней грани. Тогда элементарными событиями будут шесть событий, заключающихся в выпадении единицы, двойки, тройки, четвёрки, пятёрки и шестёрки на верхней грани кубика. Достоверным событием будет «выпадение числа очков меньше семи». Невозможным событием будет выпадение нуля очков. Случайных событий, не являющимся элементарными, достаточно много, например, таким событием является «выпадение числа очков меньше пяти».

Понятие вероятности

Что такое *вероятность*? Это слово часто используется в обыденной жизни. Им обозначают меру уверенности в том, что произойдёт некоторое событие. При этом обычно вероятность выражают в процентах, имея в виду, что если событие обязательно произойдёт, то вероятность этого равна 100%. Так, в прогнозе погоды можно услышать, что вероятность дождя в ближайшие два часа равна 60%.

Разберёмся, что имеется в обычной жизни в виду, когда утверждается, что вероятность события A равна числу p? Попытайтесь сами себе ответить на вопрос: «Что такое вероятность события 60%?» Скорее всего, Ваш ответ будет такой: «Событие произойдёт в 60 случаях из 100». Вы на верном пути. Просто слишком категорично. Неужели Вы всерьёз думаете, что событие произойдёт ровно в 60 случаях из 100? Действительно, такая предопределённость, особенно во время сотого испытания выглядит весьма неправдоподобно. Более верным следует признать такой ответ: «Событие произойдет примерно в 60 случаях из 100». Согласен. Остался другой вопрос: «Что такое примерно 60? Это больше чего и меньше чего?» Не нужно ждать строгих ответов на такие вопросы.

Чтобы окончательно разобраться в том, что такое вероятность с житейской точки зрения, введём понятие частоты появления события. Пусть произведено N испытаний и событие A произошло M раз. Тогда *частотой* f появления события будем называть дробь M/N. Предположим, что проводимый эксперимент характеризуется *устойчивостью частот*. Под этим термином подразумевается, что в разных достаточно длинных сериях испытаний, частоты появления события A не сильно отличаются друг от друга. В опытах с такими условиями обычно наблюдается, что чем больше длина серии экспериментов, тем частота появления события ближе к некоторому числу, являющемуся внутренней характеристикой события (которое и называется вероятностью). Если Вы помните тему «пределы» из курса математики, то может быть, Вам пришла мысль о том, что вероятность — это предел частоты, при стремлении числа испытаний к бесконечности. Практически идеально, только опять не надо так категорично. Если бы это было так, то согласно определению предела последовательности, для любого числа $\varepsilon>0$, существовал бы такой номер испытания, начиная с которого, обязательно выполнилось бы неравенство $p-\varepsilon < f < p+\varepsilon$. Подобного ограничения не может быть, так как теоретически (но не практически!) частота f может принимать любые, допустимые правилами арифметики, значения. Верным будет утверждение, известное в теории вероятностей под названием *Теорема Бернулли*: **для любого числа ε>0 верно, что** $\lim\limits_{N\to\infty} P(|f_N - p| < \varepsilon) = 1$**.** Здесь f_N это частота появления события A в первых N испытаниях. Такое стремление частоты к вероятности математики называют *стремлением по вероятности*. Разумеется, это не есть стремление в полном смысле этого слова. В одном из замечательнейших учебников (Вентцель Е.С. «Теория вероятностей» [4]) для такого стремления используется глагол «*приближается*». Некоторым недостатком формулировки теоремы Бернулли, выдаваемой здесь за житейский смысл вероятности, является то, что вероятность события определяется через вероятность другого события.

Если Вы согласны с близостью вероятности и частоты при большом количестве испытаний, то теперь можете окончательно сформулировать ответ на вопрос: «Что такое вероятность 60%?» Видимо ответ такой: «При большом числе испытаний частота события приближается, пусть с некоторыми временными отклонениями, к 3/5». Последний вопрос: «А как быть с вероятностью дождя в ближайшие два часа?» Действительно, не может быть и речи о проведении этого эксперимента многократно. Ближайшие два часа пройдут раз и навсегда. Дождь либо будет, что приведёт к частоте, равной единице, либо не будет, и тогда частота этого события окажется равной нулю. В любой ситуации, никакого приближения частоты к вероятности не ожидается. Чтобы решить эту проблему запомните свои ощущения относительно меры уверенности того, что в многократно повторяющемся эксперименте произойдёт событие с аналогичной вероятностью, и перенесите эти ощущения на событие в неповторяемом эксперименте.

Аксиомы вероятности

В математике даётся следующее определение понятия вероятность: каждому событию A ставится в соответствие вещественное число $P(A)$, удовлетворяющее следующим трём аксиомам:

1) *Неотрицательность*. $P(A)\geq 0$,
2) *Нормированность*. $P(\Omega)=1$,
3) *Аддитивность* (*аксиома сложения*). Если $A_i \cap A_j = \varnothing$ при $i \neq j$, то $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$, или то же самое в развёрнутом виде: $P(A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_n \cup \ldots) = P(A_1)+P(A_2)+P(A_3)+\ldots+P(A_n)+\ldots$. В частности, для двух событий: если $A \cap B = \varnothing$, то $P(A \cup B) = P(A)+P(B)$. Или то же самое словами: **Вероятность наступления хотя бы одного из попарно несовместных событий равна сумме вероятностей этих событий.**

Теперь возникает закономерный вопрос гуманитария: «Почему так, и какая связь этого с житейским пониманием вероятности». Выше было установлено, что вероятность приближённо равна частоте (при достаточно большом количестве испытаний) и «приближается» к ней при неограниченном увеличении числа испытаний (я всё-таки глагол «приближается» пишу в кавычках, так как ни о какой монотонности убывания величины $|f-p|$ не может быть и речи). И житейские представления о вероятности, кроме как с понятием частоты, связать не с чем. Убедимся в выполнении свойств частоты, соответствующих трём аксиомам вероятности.

1) Частота равна отношению двух неотрицательных чисел (числа наступивших событий и количества испытаний) и не может быть отрицательной.
2) Если событие достоверное, то оно происходит при любом испытании. Число появлений события совпадает с числом экспериментов. Оказываются равными числитель и знаменатель дроби, равной частоте. Поэтому частота появления достоверного события равна единице.
3) Разберём простейший случай. Пусть событий не бесконечно много, а всего два. Пусть поведено N испытаний, событие A произошло K раз, а событие B наступило L раз. Причём, поскольку A и B несовместные события, то одновременно они произойти не могли, и, следовательно, <u>хотя бы</u> одно из этих событий наступило столько же раз, сколько наступило <u>ровно</u> одно из них, то есть $(K+L)$ раз. Тогда частота наступления события A равна K/N, частота наступления события B равна L/N, а частота наступления хотя бы одного из этих событий равна $(K+L)/N$. И действительно справедливо равенство $K/N+L/N=(K+L)/N$.

Теперь остаётся надеяться (а математик аксиомы постулирует, не испытывая никаких психологических проблем), что у вероятности свойства такие же как и у частоты.

Следствия из аксиом вероятности

1. *Вероятность противоположного события* $P(\overline{A}) = 1 - P(A)$.
   Доказательство. Поскольку $A \cap \overline{A} = \varnothing$ и $A \cup \overline{A} = \Omega$, то по нормированности вероятности и аксиоме сложения: $1 = P(\Omega) = P(A \cup \overline{A}) = P(A) + P(\overline{A})$. Следовательно $P(\overline{A}) = 1 - P(A)$, как впрочем и $P(A) = 1 - P(\overline{A})$. Это свойство часто используется при нахождении вероятностей реализации хотя бы одного из некоторого набора событий. Обычно гораздо проще найти вероятность противоположного события (заключающееся в том, что должны будут одновременно реализоваться все события из этого набора) и вычесть её из единицы.
2. *Вероятность невозможного события* $P(\varnothing)=0$.
   Доказательство. Поскольку $\varnothing = \Omega \backslash \Omega = \overline{\Omega}$, то по вероятности противоположного события $P(\varnothing)=P(\overline{\Omega})=1 - P(\Omega)=1-1=0$.
3. *Монотонность* вероятности. Если $A \subset B$, то $P(A) \le P(B)$.
   Доказательство. Разобьём событие B на два несовместных события A и $B \backslash A$ (см. рис. 1). Теперь из $A \cap (B \backslash A)=\varnothing$ и $A \cup (B \backslash A)=B$ по аксиоме сложения следует, что $P(B)=P(A)+P(B \backslash A)$. Но по неотрицательности вероятности $P(B \backslash A) \ge 0$ и поэтому $P(B) = P(A) + P(B \backslash A) \ge P(A) + 0 = P(A)$. Это свойство названо монотонностью, поскольку напоминает определение возрастающей функции на промежутке, согласно которому большему значению аргумента соответствует большее значение функции.

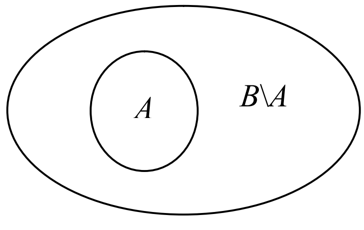

Рис. 1. Событие *B*.

4. *Ограниченность* вероятности. $P(A) \le 1$.
   Доказательство. Любое событие $A \subset \Omega$. По свойству монотонности и аксиоме нормированности $P(A) \le P(\Omega) = 1$.
5. *Правило сложения.* $P(A \cup B)=P(A)+P(B)-P(A \cap B)$. В отличие от аксиомы сложения здесь события A и B какие угодно (в аксиоме сложения они были обязательно несовместные).
   Доказательство. Разобьём событие $A \cup B$ на три несовместных события $A \backslash B$, $B \backslash A$ и $A \cap B$ (см. рис. 2). Теперь из того, что $A \cup B=(A \backslash B) \cup (B \backslash A) \cup (A \cap B)$, и попарной несовместности этих трёх событий, то есть из того, что $(A \backslash B) \cap (B \backslash A)=\varnothing$, $(A \backslash B) \cap (A \cup B)=\varnothing$ и $(B \backslash A) \cap (A \cup B)=\varnothing$, согласно аксиоме сложения следует, что $P(A \cup B)=P(A \backslash B)+P(B \backslash A)+P(A \cap B)$. Заметим, что при этом разбиении, событие A автоматически окажется разбитым на два несовместных события: $A \backslash B$ и $A \cap B$. И из того, что $(A \backslash B) \cap (A \cap B)=\varnothing$ и $A=(A \backslash B) \cup (A \cap B)$ по аксиоме сложения следует, что $P(A)=P(A \backslash B)+P(A \cap B)$. Аналогичное явление произойдёт и с событием B. Оно тоже будет разбито на два несовместных события: $B \backslash A$ и $A \cap B$. И из того, что $(B \backslash A) \cap (A \cap B)=\varnothing$ и $B=(B \backslash A) \cup (A \cap B)$ по аксиоме сложения следует, что $P(B)=P(B \backslash A)+P(A \cap B)$. Теперь из трёх полученных равенств с помощью арифметики можно доказать требуемое утверждение:

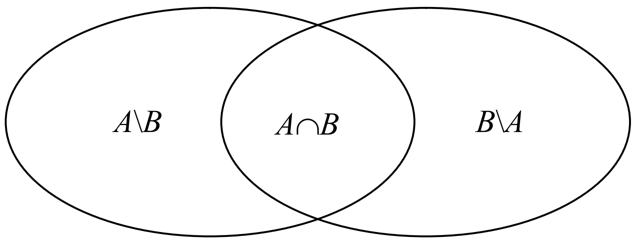

Рис. 2. Событие $A \cup B$.

$$\begin{cases} P(A \cup B) = P(A \backslash B) + P(B \backslash A) + P(A \cap B), \\ P(A) = P(A \backslash B) + P(A \cap B), \\ P(B) = P(B \backslash A) + P(A \cap B), \end{cases} \Leftrightarrow \begin{cases} P(A \cup B) = P(A \backslash B) + P(B \backslash A) + P(A \cap B), \\ P(A \backslash B) = P(A) - P(A \cap B), \\ P(B \backslash A) = P(B) - P(A \cap B). \end{cases}$$

Подставляя второе и третье равенства, в первое, получим
$P(A\cup B)=P(A)-P(A\cap B)+P(B)-P(A\cap B)+P(A\cap B)=P(A)+P(B)-P(A\cap B)$.

6. *Неравенство для совокупности событий.* $P\left(\bigcup_{i=1}^{\infty}A_i\right)\le\sum_{i=1}^{\infty}P(A_i)$, или то же самое в развёрнутом виде: $P(A_1\cup A_2\cup A_3\cup\ldots\cup A_n\cup\ldots)\le P(A_1)+P(A_2)+P(A_3)+\ldots+P(A_n)+\ldots$. В частности для двух событий: $P(A\cup B)\le P(A)+P(B)$.
Доказательство проведём только для двух событий. По правилу сложения $P(A\cup B)=P(A)+P(B)-P(A\cap B)$. По свойству неотрицательности вероятности $P(A\cap B)\ge 0$. Следовательно, $P(A\cup B)=P(A)+P(B)-P(A\cap B)\le P(A)+P(B)-0=P(A)+P(B)$. Напомним, что равенство наступает при $A\cap B=\varnothing$.

Классическое определение вероятностей

Пусть при проведении эксперимента возможно появление n различных элементарных событий. Обозначим их ω_1, ω_2, ω_3, … , ω_n. Таким образом, они оказываются попарно несовместными (при $i\neq j$ $\omega_i\cap\omega_j=\varnothing$) и единственно возможными ($\omega_1\cup\omega_2\cup\omega_3\cup\ldots\cup\omega_n=\Omega$). По аксиоме сложения $P(\omega_1)+P(\omega_2)+P(\omega_3)+\ldots+P(\omega_n)=P(\omega_1\cup\omega_2\cup\omega_3\cup\ldots\cup\omega_n)=P(\Omega)=1$. Если теперь предположить, что события ω_1, ω_2, ω_3, … , ω_n равновозможные, то их вероятности логично считать равными друг другу: $P(\omega_1)=P(\omega_2)=P(\omega_3)=\ldots=P(\omega_n)$. И поэтому сумма их вероятностей это сумма равных друг другу n слагаемых:
$1=P(\omega_1)+P(\omega_2)+P(\omega_3)+\ldots+P(\omega_n)=P(\omega_1)+P(\omega_1)+P(\omega_1)+\ldots+P(\omega_1)=n\ P(\omega_1)$. Следовательно, $P(\omega_1)=1/n$. Как впрочем, и все остальные: $P(\omega_1)=P(\omega_2)=P(\omega_3)=\ldots=$ $=P(\omega_n)=1/n$. Пусть событие A состоит из m элементарных событий, для конкретности первых m: $A=\omega_1\cup\omega_2\cup\omega_3\cup\ldots\cup\omega_m$. По аксиоме сложения вероятность события A будет равна сумме m одинаковых слагаемых:
$P(A)=P(\omega_1)+P(\omega_2)+P(\omega_3)+\ldots+P(\omega_m)=1/n+1/n+1/n+\ldots+1/n=m\cdot(1/n)=m/n$. Если события ω_1, ω_2, ω_3, … , ω_n не будут равновозможными, то нельзя их вероятности считать равными друг другу и выведенная формула не будет справедливой. Равновозможность событий ω_1, ω_2, ω_3, … , ω_n обычно следует из симметричности задачи. Например, если игральная кость сделана без смещённого центра тяжести, то нет оснований предпочесть какую-то из её граней другим. Для игральной кости $P(\omega_1)=P(\omega_2)=P(\omega_3)=P(\omega_4)=P(\omega_5)=P(\omega_6)=1/6$. Существуют и другие объекты, обладающие свойством симметрии. Простейшим из них является монета (если не оговорено, что она гнутая). При подбрасывании монеты возможны два исхода: выпадение герба и выпадение решки на её верхней грани. Для симметричной монеты $P(\omega_1)=P(\omega_2)=1/2$. Можно организовать эксперимент с любым заранее данным числом n равновозможных исходов. Для этого можно взять колоду из n карт, перемешать её и наугад выбрать одну карту. Тот же эффект достигается, если n различных шариков поместить в урну, и вытащить один из них не глядя. Подобный *классический* способ нахождения вероятностей полностью соответствует указанным выше аксиомам, и поэтому «имеет право» называться вероятностью.

Геометрическое определение вероятности

Пусть эксперимент состоит из бросания точки x на промежуток $[a;b)$, причём вероятность попадания этой точки в любые промежутки равной длины, являющиеся его частями одинаковая. Исследуем вероятности попадания точки в два промежутка различной длины. Обозначим эти длины l_1 и l_2. Пусть отношение длин этих промежутков равно отношению двух натуральных чисел: $l_1/l_2=k_1/k_2$. Тогда по свойству пропорции $l_1/k_1=l_2/k_2=m$. Здесь m имеет смысл длины k_1 непересекающихся промежутков, составляющих промежуток длины l_1 и k_2 непересекающихся промежутков, составляющих промежуток длины l_2. Пусть вероятность попадания в промежуток длины m равна p. Тогда по аксиоме сложения вероятности попадания в промежутки длин l_1 и l_2 соответственно равны pk_1 и pk_2. И отношение этих вероятностей окажется равной отношению длин промежутков: $pk_1/(pk_2)=l_1/l_2$. Логично предположить, что этим свойством обладают и промежутки, отношение длин которых нельзя выразить рациональным числом (любое иррациональное число можно представить как предел последовательности рациональных чисел), то есть вероятность попадания в промежуток при вышеупомянутом условии всегда пропорциональна его длине. Поскольку попадание точки x в промежуток $[a;b)$ событие достоверное, то вероятность этого равна единице. И, следовательно, для любого промежутка $[c;d)\subset[a;b)$ справедливо:

$$\frac{P(x\in[c;d))}{P(x\in[a;b))}=\frac{d-c}{b-a}\Leftrightarrow\frac{P(x\in[c;d))}{1}=\frac{d-c}{b-a}\Leftrightarrow P(x\in[c;d))=\frac{d-c}{b-a}.$$

Полученная формула позволяет обнаружить *парадокс нулевой вероятности* заключающийся в существовании событий, не являющихся невозможными, но тем не менее, вероятности которых равны нулю (такие события называются *почти невозможными*). Действительно, если поставить вопрос о попадании точки x в точку c (это событие возможно), которую можно трактовать как предельное состояние промежутка $[c;d)$ при d стремящимся к c, то $P(x\in\{c\})=\lim\limits_{d\to c+0}P(x\in[c;d))=\lim\limits_{d\to c+0}\frac{d-c}{b-a}=\frac{c-c}{b-a}=\frac{0}{b-a}=0$. Если развивать эту мысль дальше, то окажется, что отрезок, на который точка неизбежно упадёт, состоит только из таких точек, вероятности попадания в которые равны нулю. И, следовательно, в первом же эксперименте (как впрочем, и во всех остальных) обязательно произойдёт почти невозможное событие! По теореме Бернулли при $p=0$ для любых чисел $\varepsilon>0$ и $\delta>0$ существует номер испытания, начиная с которого $P(f<\varepsilon)>1-\delta$. Здесь частота f вовсе не обязательно равна нулю. Из этой теоремы следует, что при неограниченном увеличении числа опытов, частота почти невозможного события, опустится ниже любого, наперёд заданного значения со сколь угодно большой вероятностью. Если перейти от теоретических бесед к практике (источнику здравого смысла) и пытаться получить почти невозможное событие в опыте, то для фиксации такого события придётся обзавестись измерительным прибором бесконечной точности, что невозможно.

Естественно ожидать, что существуют и события с вероятностью равной единице, не являющиеся достоверными. Их называют *почти достоверными*. Это события, противоположные почти невозможным событиям. Например, это событие $x\in[a;c)\cup(c;b)$.

Как организовать эксперимент с бросанием точки на отрезок, именно так, чтобы вероятность попадания этой точки в любые промежутки равной длины, одинаковая? Можно предложить нанести числа от a до b на обод велосипедного колеса (при этом концы промежутка, числа a и b, геометрически будут находиться в одной точке обода, именно по этой причине один из концов промежутка в него не был включён), поднять колесо над поверхностью земли, раскрутить его, и в случайно выбранный момент нажать на тормоз. То число, которое окажется напротив тормозной колодки и определит точку x.
Геометрическим способом можно также определить вероятность попадания точки на часть плоской фигуры, при условии, что вероятности попадания этой точки в любые фрагменты фигуры равной площади одинаковые. В этом случае вероятность будет равна отношению площади части фигуры к площади всей фигуры. Подобный *геометрический* способ нахождения вероятностей полностью соответствует указанным выше аксиомам, и, также как и классический способ, тоже «имеет право» называться вероятностью.

Условная вероятность

Вероятность события A при условии, что произошло событие B, называется *условной вероятностью* и обозначается $P(A|B)$. Рассмотрим только случай, когда $P(B)>0$ (даже если событие В произошло, то в силу парадокса нулевой вероятности, ещё нет оснований полагать, что $P(B)\neq 0$). В математике даётся следующее определение понятия условной вероятности при $P(B) > 0$: $P(A|B)=\dfrac{P(A\cap B)}{P(B)}$. Это очень странное заявление для представителей гуманитарных наук. Действительно, если словесное описание дано, то как можно постулировать формулу? Она должна быть выведена из словесного описания! Или наоборот: почему эта формула равна вероятности события A при условии, что событие B произошло?

Посмотрите на рисунок 3. На нём большим прямоугольником обозначено достоверное событие Ω, а эллипсами случайные события A и B. Если принять за факт, что событие B произошло, то теперь уже оно начинает играть роль достоверного события. Для случаев, когда вероятность можно вычислить классическим (на основании равновозможности элементарных событий) или геометрическим способами, можно доказать, что отношения друг к другу вероятностей различных событий, влекущих за собой наступление события B (в частности событий $A\cap B$ и собственно B), остаются неизменными. Будем надеяться, что это свойство остаётся справедливым и для других экспериментов, вероятности событий в которых нельзя вычислить ни с помощью классического ни с помощью геометрического способов. Тогда получим: $\dfrac{P(A\cap B)}{P(B)}=\dfrac{P((A\cap B)|B)}{P(B|B)}=\dfrac{P(A|B)}{1}=P(A|B)$. В этих равенствах использовано, что при наступлении события B, говорить о наступлении события A и события $A\cap B$ то же самое, поскольку оставшаяся часть ($A\backslash B$) события A никак уже произойти не сможет, и поэтому действительно $P((A\cap B)|B)=P(A|B)$.

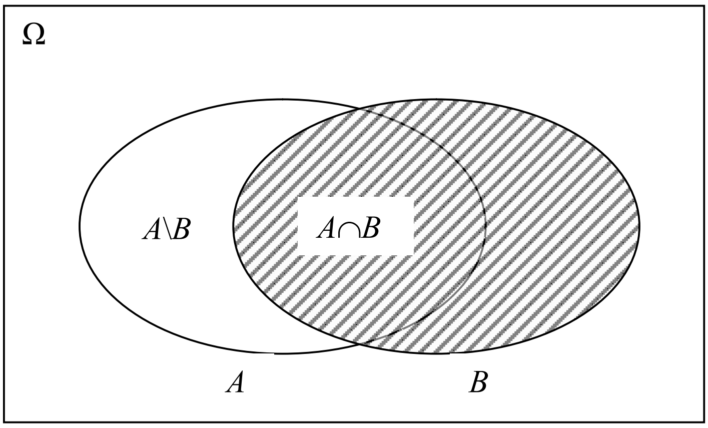

Рис. 3. Событие B произошло.

Вернёмся к случаю $P(B) = 0$. Условная вероятность $P(A|B)$ (с точки зрения математика она не определена), с точки зрения гуманитария, равна вероятности события A при условии, что произошло событие B. Напомним, что вероятность в житейском понимании определяется через частоту появления события в эксперименте при стремлении числа испытаний (которое будет равно числу появления события B) к бесконечности. Теперь вопрос: как организовать эксперимент, чтобы в нём событие В с нулевой вероятностью произошло бесконечное число раз, при отсутствии технической возможности зафиксировать такое событие?

Правило произведения

Из равенства, служащего определением условной вероятности $P(A|B)=\dfrac{P(A\cap B)}{P(B)}$ следует формула: $P(A\cap B)=P(A|B)\cdot P(B)$. Её называют *правило произведения*. Она используется для нахождения вероятности одновременной реализации двух событий. Сначала вычисляется вероятность какого-то одного из этих событий, а потом, вообразив, что первое событие уже произошло, находят вероятность второго события. Перемножая эти две вероятности, получают окончательный результат.

Формула полной вероятности

Рассмотрим двухступенчатый эксперимент, в котором на первой стадии обязательно происходит ровно одно из n событий: H_1, H_2, H_3, … , H_n (буква H это первая буква в слове «гипотеза»). Для этого эти n событий должны быть попарно несовместными (при $i \neq j$ $H_i \cap H_j = \varnothing$) и единственно возможными ($H_1 \cup H_2 \cup H_3 \cup \ldots \cup H_n = \Omega$) событиями. На второй стадии эксперимента, после того как уже одно из событий H_i произошло, происходит или не происходит событие A. Схема опыта изображена на рисунке 4.

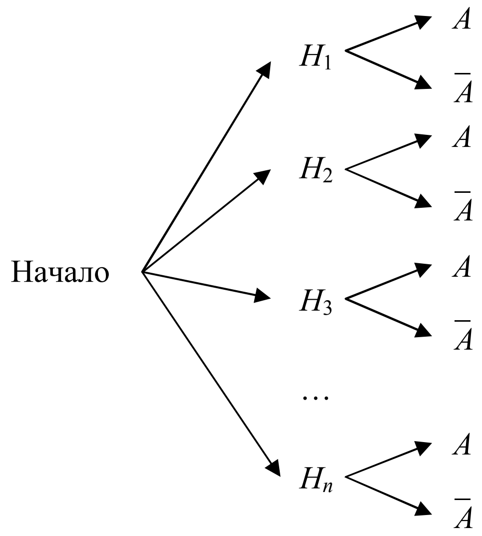

Рис. 4. Двухступенчатый эксперимент.

Представим событие A в виде объединения непересекающихся событий: $A = A \cap \Omega = A \cap (H_1 \cup H_2 \cup H_3 \cup \ldots \cup H_n) = (A \cap H_1) \cup \cup (A \cap H_2) \cup (A \cap H_3) \cup \ldots \cup (A \cap H_n)$ (на рисунке 5 показано разложение события A, обозначенного эллипсом). По аксиоме сложения вероятностей имеем: $P(A) = = P(A \cap H_1) + P(A \cap H_2) + P(A \cap H_3) + \ldots + + P(A \cap H_n)$.

Если все события H_i имеют ненулевые вероятности (иначе не записать условные вероятности), то используя правило произведения для каждого слагаемого, получим: $P(A) = P(A|H_1) \cdot P(H_1) + P(A|H_2) \cdot P(H_2) + + P(A|H_3) \cdot P(H_3) + \ldots + P(A|H_n) \cdot P(H_n)$. Выведенная формула называется *формулой полной вероятности*.

Формула гипотез (Байеса)

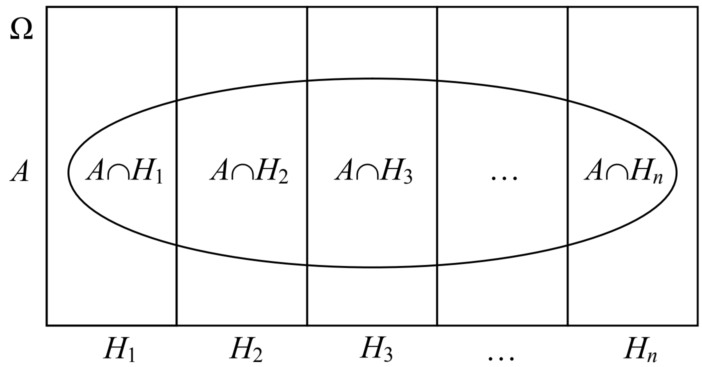

Рис. 5. Разложение события A.

Пусть завершился двухступенчатый эксперимент, описанный в предыдущем пункте. Пусть на его второй стадии произошло событие A. Предположим, что $P(A) \neq 0$ (иначе не записать условную вероятность). Поставим вопрос: какова условная вероятность того, что на первой стадии эксперимента произошло событие H_k (где $k=1, 2, 3, \ldots, n$)? Эта условная вероятность $P(H_k|A)$ называется *апостериорной*, в отличие от вероятности $P(H_k)$, которая называется *априорной*. Эти две вероятности одного и того же события могут отличаться друг от друга, поскольку априорная вероятность вычисляется до эксперимента, а апостериорная после, когда уже появилась дополнительная информация, используя которую возможно скорректировать вероятность события H_k. По формуле, являющейся определением условной вероятности, имеем: $P(H_k \mid A) = \dfrac{P(H_k \cap A)}{P(A)}$. Теперь числитель этой дроби преобразуем по правилу произведения, а знаменатель по формуле полной вероятности:

$$P(H_k \mid A) = \frac{P(A \mid H_k)P(H_k)}{P(A \mid H_1)P(H_1) + P(A \mid H_2)P(H_2) + P(A \mid H_3)P(H_3) + \ldots + P(A \mid H_n)P(H_n)}.$$

Выведенная формула называется *формулой Байеса*.

Независимость двух событий

В математике даётся следующее определение *независимости* событий: «два события A и B называются независимыми, если $P(A \cap B)=P(A) \cdot P(B)$».

Что такое, с точки зрения здравого смысла, независимость события A от события B? Это означает, что вероятность события A никак не изменится, если произойдёт событие B. На языке формул это выглядит следующим образом: $P(A)=P(A|B)$. В записи этого соотношения использована условная вероятность, поэтому придётся предположить, что $P(B) \neq 0$.

Используя формулу (при $P(B) \neq 0$), служащую определением условной вероятности получим: $P(A) = P(A \mid B) = \dfrac{P(A \cap B)}{P(B)} \Leftrightarrow P(A)P(B) = P(A \cap B)$. Действительно, получилось соотношение, которое было дано в определении. Интересно, что в него события A и B входят симметрично. Дело в том, что произведение не меняется, если переставить местами сомножители. Аналогичным свойством обладает и фраза об одновременном появлении событий. Если предположить, что и $P(A) \neq 0$, то по правилу произведения можно записать: $P(A) \cdot P(B)=P(A \cap B)=P(B \cap A)=P(B|A) \cdot P(A)$. Сокращая это равенство на $P(A) \neq 0$, получим: $P(B)=P(B|A)$. То есть, вероятность события B тоже не изменяется в зависимости от того, произошло ли событие A. Таким образом, если $P(A) \neq 0$ и $P(B) \neq 0$, то независимость событий A и B друг от друга может быть только двусторонней: если событие A не зависит от события B, то и событие B не зависит от события A.

Заметим, что определение независимости событий формально можно применять и для событий с нулевой вероятностью, но при этом будет невозможно сравнить результат с соображениями здравого смысла, поскольку они оперируют понятием условной вероятности.

Свойства независимых событий

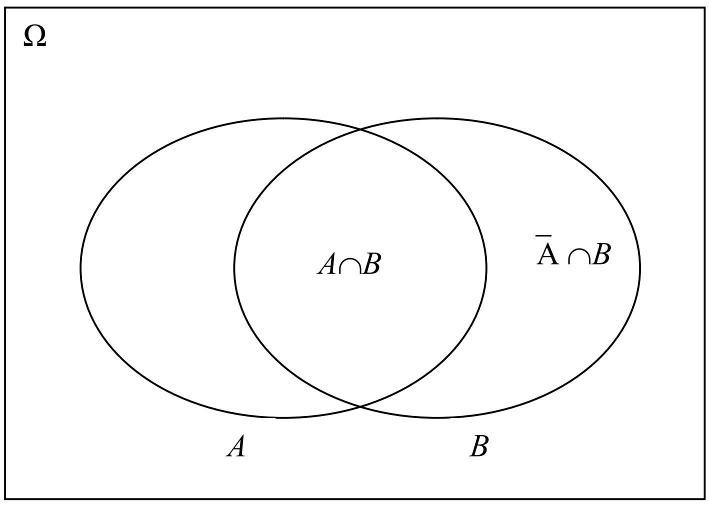

Рис. 6. Разбиение события B.

1. Если события A и B независимые, то и события A и $\overline{B}$ тоже независимые.
   Доказательство. Представим событие B в виде объединения несовместных событий (см. рис. 6): $B=\Omega \cap B=(A \cup \overline{A}) \cap B=(A \cap B) \cup$ $\cup (\overline{A} \cap B)$. Тогда по аксиоме сложения $P(B)=P(A \cap B)+P(\overline{A} \cap B)$. Из независимости событий A и B следует, что $P(A \cap B)=P(A) \cdot P(B)$. Подставляя это равенство в предыдущее, учитывая свойство вероятности противоположного события, получим: $P(B)=P(A) \cdot P(B)+\mathrm{P}(\overline{A} \cap B) \Leftrightarrow P(B)-P(A) \cdot P(B)=P(\overline{A} \cap B) \Leftrightarrow$ $\Leftrightarrow P(B) \cdot (1-P(A))=P(\overline{A} \cap B) \Leftrightarrow P(B) \cdot P(\overline{A})=P(\overline{A} \cap B)$. Последнее равенство и доказывает независимость событий A и $\overline{B}$.
2. Если $P(B)=0$, то событие B независимо с любым событием.
   Доказательство. Пусть вероятность события B равна нулю: $P(B)=0$. Для этого оно должно быть либо невозможным, либо почти невозможным. Заметим, что $(A \cap B) \subset B$

и по аксиоме неотрицательности и свойству монотонности вероятности: $0 \le P(A \cap B) \le P(B)=0$, и, следовательно, $P(A \cap B)=0$. Тогда $P(A \cap B)=P(A) \cdot P(B) \Leftrightarrow$ $\Leftrightarrow 0=P(A) \cdot 0 \Leftrightarrow 0=0$ верное равенство, что и доказывает независимость событий A и B. С философской точки зрения так и должно быть: невозможное событие никогда не происходит, и вполне естественно, что оно ни от чего не может зависеть и от него ничего не может зависеть.

3. Если $P(B)=1$, то событие B независимо с любым событием.

   Доказательство. Пусть вероятность события B равна единице: $P(B)=1$. Для этого оно должно быть либо достоверным, либо почти достоверным. По свойству вероятности противоположного события $P(\bar{B})=1-P(B)=1-1=0$. Тогда по предыдущему свойству события A и $\bar{B}$ независимые. А поскольку противоположным к событию $\bar{B}$ является событие B, то по первому из этих свойств, события A и B независимые. Действительно достоверное событие происходит при любом испытании независимо ни от чего. Естественно, что и от него ничего не может зависеть.
4. Если события с ненулевыми вероятностями независимые, то они совместные.

   Доказательство. Пусть события A и B независимые, то есть $P(A \cap B)=P(A) \cdot P(B)$, а по условию $P(A)>0$ и $P(B)>0$. Поэтому $P(A \cap B)=P(A) \cdot P(B)>0$, и, следовательно, событие $A \cap B$ не может быть невозможным (иначе бы его вероятность была бы равна нулю). И, следовательно, события A и B обязательно совместные.
5. Если события с ненулевыми вероятностями несовместные, то они зависимые.

   Доказательство. Пусть события A и B несовместные, то есть $A \cap B=\varnothing$. Тогда $P(A \cap B)=0$. По условию $P(A)>0$ и $P(B)>0$. Следовательно, $P(A) \cdot P(B)>0$, и поэтому равенство $P(A \cap B)=P(A) \cdot P(B)$ невозможно, то есть события A и B не являются независимыми.
6. Если события с ненулевыми вероятностями противоположные, то они зависимые.

   Доказательство. Если события противоположные, то они обязательно несовместные и это свойство следует из предыдущего.

Независимость трех и более событий

Если событий три или больше, то различают два вида независимости: *попарную* и *взаимную*.

События $A_1, A_2, A_3, \ldots, A_n$ называются попарно независимыми, если независимы события в каждой из пар этих событий, то есть для любых $i \ne j$ верно, что $P(A_i \cap A_j)=P(A_i) \cdot P(A_j)$. В частности, три события A, B и C будут попарно независимыми, если $\begin{cases} P(A \cap B) = P(A) \cdot P(B), \\ P(A \cap C) = P(A) \cdot P(C), \\ P(B \cap C) = P(B) \cdot P(C). \end{cases}$

События $A_1, A_2, A_3, \ldots, A_n$ называются взаимно независимыми (или независимыми в совокупности), если для любого набора из этих n событий верно, что вероятность их одновременного появления равна произведению их вероятностей. В частности, три события A, B и C будут взаимно независимыми, если $\begin{cases} P(A \cap B) = P(A) \cdot P(B), \\ P(A \cap C) = P(A) \cdot P(C), \\ P(B \cap C) = P(B) \cdot P(C), \\ P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C). \end{cases}$

Из этих двух определений следует, что если события взаимно независимые, то они обязательно попарно независимые. Действительно, если что-то верно для любого набора, то это верно, в том числе, и для любой пары. Поэтому, когда употребляют термин «независимые события» без уточнения как именно независимые, то следует понимать, что они взаимно независимые. Разберём, что такое взаимная независимость событий с точки зрения здравого смысла на примере трёх событий. Если событие A не зависит от появления любой комбинации событий B и C, то кроме независимости от событий B и C это ещё означает независимость от события $B\cap C$. Другими словами: должно ещё выполняться равенство $P(A)=P(A|B\cap C)$. Учитывая определение условной вероятности, получим $P(A)=P(A\mid B\cap C)=\dfrac{P(A\cap(B\cap C))}{P(B\cap C)}=\dfrac{P(A\cap B\cap C)}{P(B\cap C)}\Leftrightarrow$ $P(A\cap B\cap C)=$ $=P(A)\cdot P(B\cap C)=P(A)\cdot P(B)\cdot P(C)$. Действительно, данное определение совпадает с представлениями о независимости, следующими из здравого смысла.

Рассмотрим ещё один вопрос: если из взаимной независимости следует попарная, то не следует ли из попарной взаимная? Правильный ответ: нет, не следует. Чтобы доказать это достаточно привести хотя бы один пример трёх попарно независимых событий, которые не являются взаимно независимыми. Такой пример был найден советским математиком Бернштейном. Он предложил взять правильный тетраэдр (это пространственная фигура, состоящая из четырёх одинаковых равносторонних треугольников) и покрасить одну его сторону в красный цвет, другую в синий, третью в зелёный, а на четвёртую частично разукрасить в красный цвет, частично в синий и частично в зелёный (см. рис. 7).

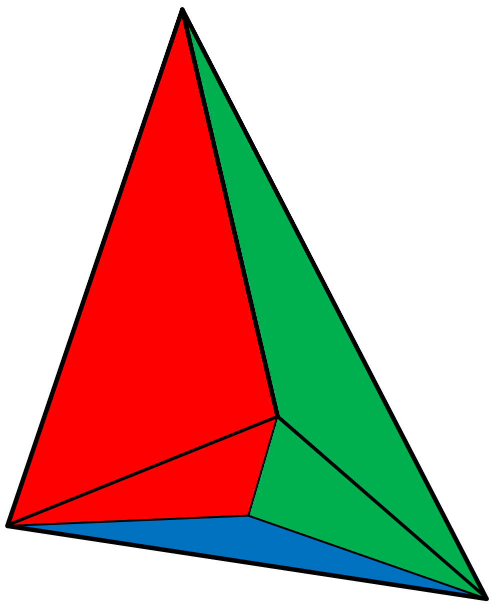

Рис. 7. Пример Бернштейна.

Экспериментом будет подбрасывание тетраэдра с фиксацией нижней грани после падения. Пусть событие A: «на оказавшейся внизу грани присутствует красный цвет», событие B: «на оказавшейся внизу грани присутствует синий цвет» и событие C: «на оказавшейся внизу грани присутствует зелёный цвет». Тогда событиями $A\cap B$, $A\cap C$, $B\cap C$ и $A\cap B\cap C$ будет событие: «на оказавшейся внизу грани присутствуют все три цвета». В силу симметрии тетраэдра, возможно использование классического способа вычисления вероятностей, согласно которому вероятности того, что тетраэдр упадёт на каждую из четырех граней, равны друг другу и равны ¼. Для реализации каждого из событий A, B и C возможно два способа: тетраэдр упадёт на грань, целиком закрашенную в соответствующий цвет, и на грань, содержащую все три цвета. Поэтому $P(A)=P(B)=P(C)=2/4=1/2$. А вероятности одновременной реализации любой комбинации этих трёх событий равны: $P(A\cap B)=P(A\cap C)=P(B\cap C)=P(A\cap B\cap C)=1/4$. Очевидно, что выполняются три равенства, обеспечивающие попарную независимость событий A, B и C:
$\begin{cases} P(A\cap B)=P(A)\cdot P(B), \\ P(A\cap C)=P(A)\cdot P(C), \\ P(B\cap C)=P(B)\cdot P(C), \end{cases} \Leftrightarrow \left\{\dfrac{1}{4}=\dfrac{1}{2}\cdot\dfrac{1}{2}\right.$. А четвёртое равенство, которое как раз и позволяет отличить взаимную независимость от попарной не выполнятся: $P(A\cap B\cap C)=P(A)\cdot P(B)\cdot P(C)\Leftrightarrow\left\{\dfrac{1}{4}=\dfrac{1}{2}\cdot\dfrac{1}{2}\cdot\dfrac{1}{2}\right.$, поскольку $\dfrac{1}{4}\neq\dfrac{1}{8}$. Поэтому введённые события A, B и C попарно независимые, но не взаимно независимые. Если интуиция подсказывала Вам, что события A, B и C не должны были быть даже попарно независимыми, то могу сказать, что интуиция Вас не подвела. Попарная независимость таких событий A, B и C не более чем арифметический фокус, и если бы бросаемая фигура не была бы симметричной, никакой независимости бы не было.

Последовательность независимых испытаний с двумя исходами

Рассмотрим эксперимент, заключающийся в подбрасывании несимметричной монеты. У несимметричной монеты (в отличие от симметричной) вероятности появления герба и решки не обязательно равны. Обозначим вероятность выпадения герба p, а решки q. Поскольку эти события несовместны, и монета обязательно упадёт либо на герб, либо на решку, то справедливо равенство $p+q=1 \Leftrightarrow q=1-p$. С точки зрения здравого смысла, рассматриваемые события независимые, то есть результат каждого подбрасывания монеты никак не зависит от результатов других подбрасываний в любых комбинациях. Выводы, полученные в этом пункте, окажутся справедливыми и для других экспериментов с двумя исходами, характеризующимися неизменностью вероятностей событий и взаимной независимостью испытаний. Исторически сложилось, что подобная задача родилась при составлении математической модели азартной игры. Это привело к появлению специфической терминологии. Событие, вероятность которого равна p называется *успехом*, а противоположное ему событие *неудачей*.

Обычно задача ставится так: «какова вероятность того, что произойдёт ровно m успехов в серии из n независимых испытаний?» Эту величину обозначают $P_n(m)$. Для решения этой проблемы ответим на пару вопросов:

1. Сколько всего существует элементарных событий, которые соответствуют появлению m успехов в n испытаниях?
   На этот вопрос отвечает раздел математики, называемый *комбинаторика*. Доказано, что количество способов выбрать m элементов из n без учёта порядка выбора (это как раз нужная ситуация: выбираются номера испытаний, в которых будут зафиксированы успехи; при этом порядок выбора не важен) равно числу сочетаний, которое обозначается C_n^m, и может быть найдено по формуле: $C_n^m = \dfrac{n!}{m!(n-m)!}$.
2. Какова вероятность каждого из таких элементарных событий?
   Поскольку успехов должно быть m (вероятность появления каждого равна p), а число испытаний n, то число неудач должно быть равно $(n-m)$, при этом вероятность неудачи известна и равна q. По правилу произведения для взаимно независимых событий получим произведение m сомножителей, равных p и $(n-m)$ сомножителей, равных q, то есть: $p^m q^{n-m}$.

Теперь искомую вероятность $P_n(m)$ найдём по аксиоме сложения, складывая вероятности интересующих элементарных событий. Все слагаемые будут равны друг другу и равны $p^m q^{n-m}$, а число слагаемых будет равно C_n^m. Поэтому $P_n(m) = C_n^m p^m q^{n-m}$. Выведенная формула называется *формулой Бернулли* (не путайте с уже упоминавшейся теоремой Бернулли).

Локальная теорема Муавра – Лапласа

Формула Бернулли, выведенная в предыдущем пункте, может оказаться практически бесполезной при значительных n. Проблема в том, что для нахождения числа сочетаний $C_n^m = \dfrac{n!}{m!(n-m)!}$ потребуется вычисление факториалов больших чисел, что затруднительно. Из локальной теоремы Муавра – Лапласа (формулировать саму теорему нет необходимости) следует, что при больших n, вероятность иметь ровно m успехов в серии из n независимых испытаний может быть приближённо найдена по формуле:

$P_n(m) \approx \frac{1}{\sqrt{npq}} \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{m-np}{\sqrt{npq}} \right)^2 \right)$. Степень точности этой формулы конечно зависит от n (чем n больше, тем точнее, а при маленьких n вполне можно пользоваться и точной формулой Бернулли) и от вероятностей p и q (чем эти вероятности ближе друг к другу и к ½, тем точнее). Для удобства работы вводится функция $\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2} \right)$ и тогда $P_n(m) \approx \frac{1}{\sqrt{npq}} \cdot \varphi\left(\frac{m-np}{\sqrt{npq}} \right)$. Функцию $\varphi(x)$ можно вычислить на любом инженерном микрокалькуляторе. График функции $\varphi(x)$ показан на рисунке 8. Функция четная, положительная, имеет при $x = 0$ максимум, совпадающий с наибольшим значением. Оно равно $y_{\max} = \varphi(0) = \frac{1}{\sqrt{2\pi}} \approx 0{,}398942$. График имеет две точки перегиба при $x = -1$ и $x = 1$. Значение функции в точках перегиба равно $y_{перегиба} = \varphi(-1) = \varphi(1) = \frac{1}{\sqrt{2\pi\, e}} \approx 0{,}241971$. Горизонтальная ось является асимптотой графика функции $\varphi(x)$. При $x<-4$ и $x>4$ график функции $\varphi(x)$ практически неотличим от своей асимптоты. Интересно, что $\int\limits_{-\infty}^{+\infty} \varphi(x)dx = 1$ (это является следствием известного интеграла Эйлера–Пуассона).

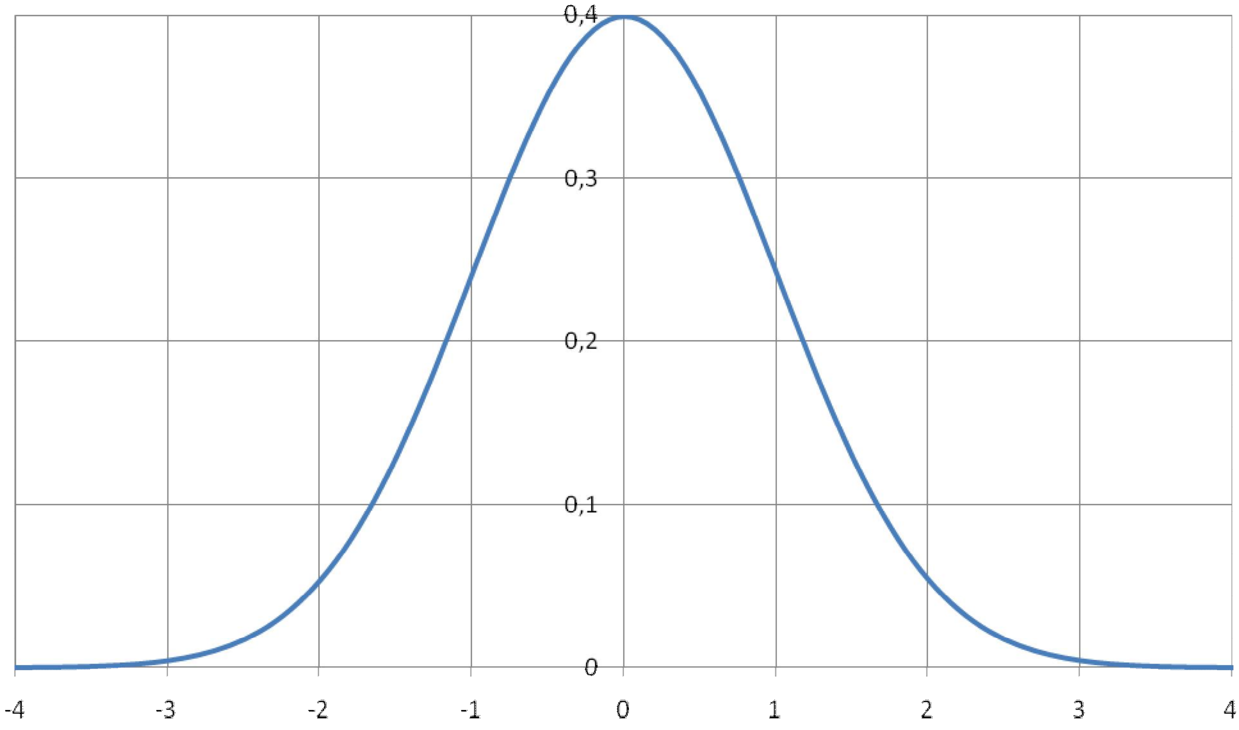

Рис. 8. График функции $y=\varphi(x)$.

Приближение Пуассона

Докажем, что если $\lambda=np$ постоянное число, то $\lim\limits_{n\to\infty} P_n(m) = \frac{\lambda^m}{m!}\cdot e^{-\lambda}$. Поскольку $\lambda=np$, то $p=\frac{\lambda}{n}$, $P_n(m) = C_n^m p^m q^{n-m} = \frac{n!}{m!(n-m)!}\left(\frac{\lambda}{n}\right)^m (1-p)^{n-m}$. Сократим дробь $\frac{n!}{(n-m)!} = (n-m+1)(n-m+2)...(n-2)(n-1)n$,

получив произведение m сомножителей. Тогда

$P_n(m) = \frac{n-m+1}{n}\cdot\frac{n-m+2}{n}\cdot ...\cdot\frac{n-2}{n}\cdot\frac{n-1}{n}\cdot\frac{n}{n}\cdot\frac{\lambda^m}{m!}\cdot\left(1-\frac{\lambda}{n}\right)^{n-m}$. Предел каждой из первых m дробей равен единице: $\lim\limits_{n\to\infty}\frac{n-m+k}{n} = \lim\limits_{n\to\infty}\left(1-\frac{m-k}{n}\right)=1$, для любого k, а предел последнего сомножителя:

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{n-m} = \lim_{n\to\infty}\exp\ln\left(1-\frac{\lambda}{n}\right)^{n-m} = \exp\lim_{n\to\infty}(n-m)\ln\left(1-\frac{\lambda}{n}\right) = \exp\lim_{n\to\infty}(n-m)\left(-\frac{\lambda}{n}\right) =$$

$= \exp\lim\limits_{n\to\infty}\left(-\lambda\left(1-\frac{m}{n}\right)\right) = \exp(-\lambda) = e^{-\lambda}$. При вычислении этого предела $\ln\left(1-\frac{\lambda}{n}\right)$ был заменён на эквивалентную бесконечно малую функцию $\left(-\frac{\lambda}{n}\right)$. Оставшаяся дробь $\frac{\lambda^m}{m!}$ по условию теоремы константа. В итоге $\lim\limits_{n\to\infty} P_n(m) = 1\cdot 1\cdot 1\cdot ...\cdot 1\cdot 1\cdot\frac{\lambda^m}{m!}\cdot e^{-\lambda} = \frac{\lambda^m}{m!}\cdot e^{-\lambda}$.

Упоминалось, что точность формулы, следующей из локальной теоремы Муавра – Лапласа зависит от того, насколько близки p и q друг к другу и к ½. Из доказанной сейчас теоремы следует другая приближённая формула, более точная в случае, когда p мало: $P_n(m) \approx \frac{(np)^m}{m!}\cdot e^{-np}$. Эту формулу рекомендуется использовать для нахождения вероятностей редких событий.

Интегральная теорема Муавра – Лапласа

Поставим вопрос: какова вероятность того, что число успехов m в серии из n независимых испытаниях будет удовлетворять неравенству $a\le m<b$? Разумеется $P(a\le m<b)=P_n(a)+P_n(a+1)+P_n(a+2)+P_n(a+3)+\ldots+P_n(b-3)+P_n(b-2)+P_n(b-1)$. Каждое из этих слагаемых может быть вычислено точно по формуле Бернулли, или приближённо (при больших n) с использованием локальной теоремы Муавра–Лапласа или приближения Пуассона. Если слагаемых слишком много, то нахождение этой суммы может оказаться затруднительным. Из интегральной теоремы Муавра – Лапласа (формулировать саму теорему нет необходимости) следует, что при больших n, вероятность того, что число успехов m в серии из n независимых испытаниях будет удовлетворять неравенству $m_1\le m<m_2$ может быть приближённо найдена по формуле: $P(a\le m<b)\approx\Phi\left(\frac{b-np}{\sqrt{npq}}\right)-\Phi\left(\frac{a-np}{\sqrt{npq}}\right)$. Здесь и в дальнейшем функцией $\Phi(x)$ обозначается такая первообразная функции $\varphi(x)$, что $\Phi(x)=\int\limits_{-\infty}^{x}\varphi(t)dt$. К сожалению эта первообразная не выражается через элементарные функции. Тем не менее, производная от функции $\Phi(x)$ известна (это функция $\varphi(x)$), и функцию $\Phi(x)$ можно исследовать обычным способом с

построением графика. Функция $\Phi(x)$ возрастает на всей числовой оси, будет иметь точку перегиба при x=0 в которой $\Phi(0)$=1/2, иметь две горизонтальные асимптоты: y=0 при $x\to-\infty$ и y=1 при $x\to+\infty$. При $x<-4$ и $x>4$ график функции $\Phi(x)$ практически неотличим от своих асимптот. График функции $\Phi(x)$ центрально симметричен относительно своей точки перегиба (см. рис. 9), что с арифметической точки зрения выражается равенством $\Phi(x)+\Phi(-x)=1$.

Значение неэлементарной функции $\Phi(x)$ в произвольной точке можно получить, использовав популярную программу Microsoft Office Excel, в русскоязычной версии которой, встроена функция НОРМСТРАСП(число). Единственный аргумент этой функции совпадает с аргументом функции $\Phi(x)$. Например, для получения значения $\Phi(0{,}58)$ в произвольную ячейку Excel нужно ввести: =НОРМСТРАСП(0,58). Результат должен быть равен 0,71904.

При создании собственной прикладной программы, например на алгоритмическом языке Delphi, функции Excel окажутся недоступными. Поэтому нужно самостоятельно описать процедуру, вычисляющую значение функции $\Phi(x)$. Для этого удобно представить $\Phi(x)$ в виде ряда: $\Phi(x)=\frac{1}{2}+\int\limits_0^x\varphi(t)dt=\frac{1}{2}+\frac{1}{\sqrt{2\pi}}\int\limits_0^x\exp\left(-\frac{t^2}{2}\right)dt=\frac{1}{2}+\frac{1}{\sqrt{2\pi}}\int\limits_0^x\sum\limits_{k=0}^{\infty}\frac{\left(-\frac{t^2}{2}\right)^k}{k!}dt=$

$$=\frac{1}{2}+\frac{1}{\sqrt{2\pi}}\int\limits_0^x\sum_{k=0}^{\infty}\frac{(-1)^k t^{2k}}{2^k k!}dt=\frac{1}{2}+\frac{1}{\sqrt{2\pi}}\sum_{k=0}^{\infty}\frac{(-1)^k}{2^k k!}\int\limits_0^x t^{2k}dt=\frac{1}{2}+\frac{1}{\sqrt{2\pi}}\sum_{k=0}^{\infty}\frac{(-1)^k}{2^k k!}\cdot\left.\frac{t^{2k+1}}{2k+1}\right|_0^x=$$

$=\frac{1}{2}+\frac{1}{\sqrt{2\pi}}\sum\limits_{k=0}^{\infty}\frac{(-1)^k x^{2k+1}}{(2k+1)2^k k!}$. Ряд получился знакочередующимся, а это по соответствующей теореме (теореме Лейбница) обеспечивает уверенность в том, что точность расчёта будет меньше модуля первого отброшенного члена. Поэтому для приближённых расчётов бесконечную сумму нужно заменить конечной, прерывая вычисления, если очередное слагаемое оказывается меньше требуемой точности.

Степень точности приближённой формулы, следующей из интегральной теоремы Муавра – Лапласа ниже, чем у двух предыдущих. Аналогично формуле, следующей из локальной теоремы Муавра – Лапласа, точность зависит от n (чем n больше, тем точнее) и от вероятностей p и q (чем эти вероятности ближе друг к другу и к ½, тем точнее). Как поступать, если p мало? К сожалению, приближение Пуассона не имеет интегральной версии. В такой ситуации исследователя ждёт неприятный выбор: либо ничего не сосчитать (или считать недопустимо долго, что иногда то же самое, что не сосчитать), либо сосчитать с низкой степенью точности.

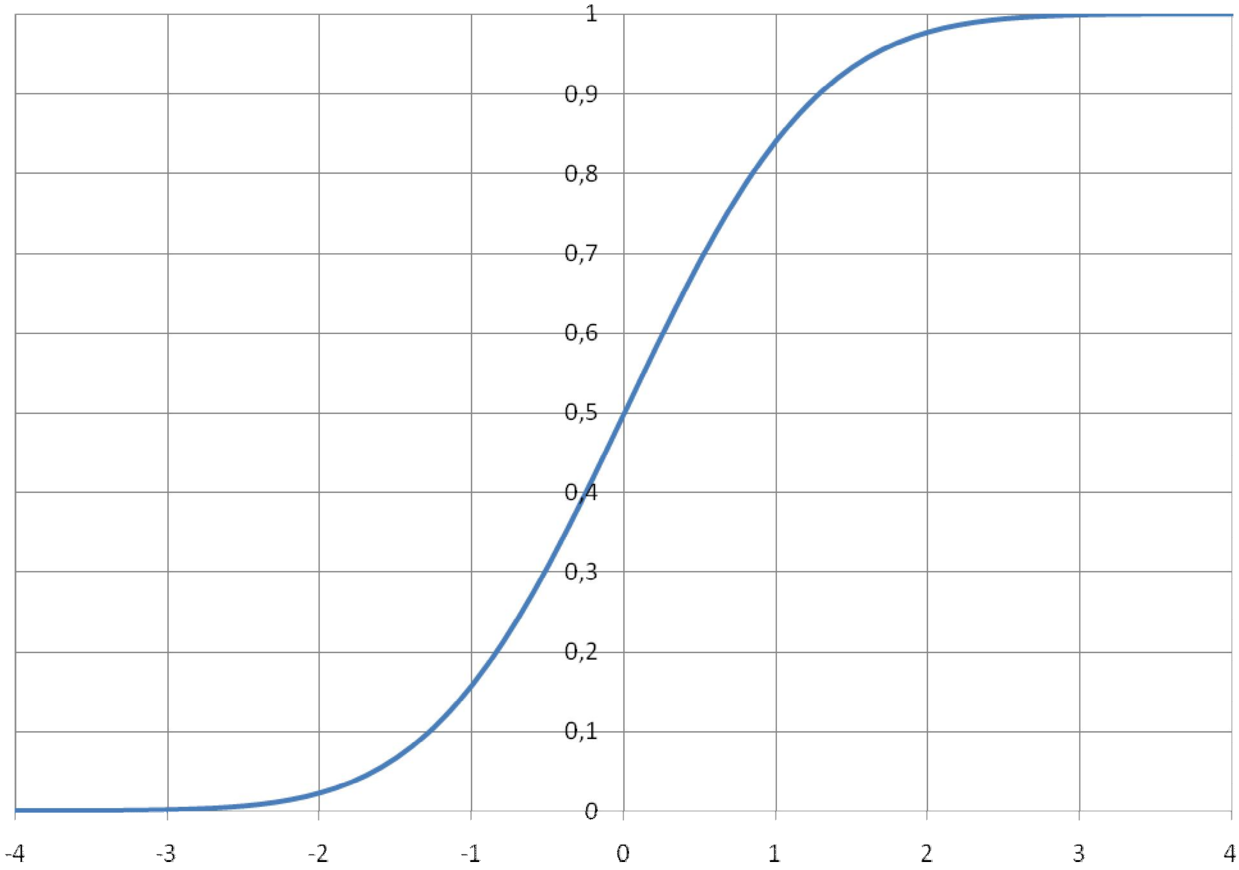

Рис. 9. График функции y=Ф(x).

С помошью интегральной теоремы Муавра – Лапласа можно попытаться ответить на вопрос, не имеющий ответа: «сколько нужно провести испытаний, чтобы гарантировать наступление хотя бы a успехов?» Естественно, что самым правильным ответом будет бесконечное число испытаний. Но, тем не менее, если 7-разрядный калькулятор не позволяет отличить $\Phi(-5)$ от нуля, то для того, чтобы $P_n(m \geq a)=1$, согласно интегральной теореме Муавра – Лапласа должно быть

$$P_n(a \leq m < +\infty) = 1 \Leftrightarrow \Phi(+\infty) - \Phi\left(\frac{a-np}{\sqrt{npq}}\right) = 1 \Leftrightarrow 1 - \Phi\left(\frac{a-np}{\sqrt{npq}}\right) = 1 \Leftrightarrow \Phi\left(\frac{a-np}{\sqrt{npq}}\right) = 0 \Leftrightarrow$$

А для этого с практической точки зрения достаточно выполнение неравенства

$$\Leftrightarrow \frac{a-np}{\sqrt{npq}} < -5 \Leftrightarrow a - np < -5\sqrt{npq} \Leftrightarrow np - 5\sqrt{npq} - a > 0 \Leftrightarrow$$

Это квадратное неравенство относительно корня из n и может быть решено заменой переменной $t = \sqrt{n}$

$$\Leftrightarrow pt^2 - 5\sqrt{pq} \cdot t - a > 0 \Leftrightarrow$$

$$D = 25pq + 4pa, \quad t_{1,2} = \frac{4\sqrt{pq} \pm \sqrt{25pq + 4pa}}{2p} = \frac{4\sqrt{q} \pm \sqrt{25q + 4a}}{2\sqrt{p}}$$

$$\Leftrightarrow \left[\begin{array}{l} t < \dfrac{4\sqrt{q} - \sqrt{25q + 4a}}{2\sqrt{p}} \\ t > \dfrac{4\sqrt{q} + \sqrt{25q + 4a}}{2\sqrt{p}} \end{array}\right. \Leftrightarrow t > \frac{4\sqrt{q} + \sqrt{25q + 4a}}{2\sqrt{p}} \Leftrightarrow \sqrt{n} > \frac{4\sqrt{q} + \sqrt{25q + 4a}}{2\sqrt{p}} \Leftrightarrow$$

$$\Leftrightarrow n > \left(\frac{4\sqrt{q} + \sqrt{25q + 4a}}{2\sqrt{p}}\right)^2 .$$

Попытаемся с помощью этой формулы ответить на вопрос: «сколько раз нужно подбросить симметричную монету, чтобы хотя бы один раз выпал герб?». В этом случае p=q=½, a=1 и тогда $n > \left(2 + \sqrt{\frac{33}{2}}\right)^2 \approx 23{,}7 \Leftrightarrow n \geq 24$. Конечно, критики скажут, что при 24 подбрасываниях возможно получить все 24 решки, причём даже известно с какой вероятностью: $\frac{1}{2^{24}} = \frac{1}{16777216}$. Но это число не поместится в разрядную сетку микрокалькулятора.

С помощью выведенной формулы построим таблицу зависимости наименьшего допустимого числа подбрасываний n, необходимого для того, чтобы симметричная монета «гарантированно» упала хотя бы a раз гербом вверх.

a	Наименьшее n
1	24
10	51
100	268
1000	2190
10000	20577
100000	201800
1000000	2005668

В таблице видно, что с ростом необходимого числа гербов a, наименьшее значение подбрасываний n растёт, асимптотически приближаясь при $a \to \infty$ к величине $\frac{a}{p} = 2a$.

Теория случайных величин

Современная теория вероятностей начинается с введения понятия *случайная величина*. До этого теория вероятностей предпочитала работать со случайными событиями.

Дискретные случайные величины

В математике есть понятие *функции*: это такое однозначное отображение, которое устанавливает соответствие между числовыми множествами. Расширим это понятие, поставив вещественное число в соответствие случаю. Будем случайной величиной называть функцию, заданную на пространстве элементарных событий. В большинстве примеров для этого практически и делать то ничего не надо. В эксперименте с бросанием игральной кости событиям: выпало одно очко, два, и так далее, достаточно поставить в соответствие какие-нибудь вещественные числа, проще всего: 1, 2, 3, 4, 5 и 6. Если эти числа почему-то не понравились, то можно на грани кубика наклеить бумажки и написать на них <u>какие угодно вещественные числа</u>. Теперь по завершению испытания будут появляться не только события, но и числа.

Можно предложить и другие (в том числе компьютерные) эксперименты, в ходе которых будут случайным образом появляться вещественные числа. Если эти числа (они называются *возможными значениями* случайной величины) стоят отдельно друг от друга на числовой прямой, то такая случайная величина называется *дискретной*. Возможных значений может быть и бесконечное количество. Каждому возможному значению должна быть поставлена в соответствие вероятность того, что случайная величина примет это значение. Традиционно обозначения при этом такие: случайные величины обозначают большими латинскими буквами из конца алфавита: например, X; возможные значения обозначают соответствующими им маленькими латинскими буквами с индексами: например, x_i; а вероятность того, что случайная величина примет значение x_i обозначается буквой p_i. В виде формулы это можно записать так: $P(X=x_i)=p_i$. Возможные значения и их вероятности обычно записывают в виде таблицы, которая называется *таблицей распределения*:

x_i:	x_1	x_2	x_3	…	x_n	…
p_i:	p_1	p_2	p_3	…	p_n	…

Возможные значения x_i могут быть любыми вещественными числами, вероятности p_i (как и все остальные вероятности) должны удовлетворять условию: $0 \le p_i \le 1$. Можно считать, что $0 < p_i \le 1$, поскольку равенство $p_i=0$ эквивалентно отсутствию в таблице распределения возможного значения x_i. Из того, что какое-то из событий $X=x_i$ обязательно должно произойти, следует необходимость выполнения *условия нормированности*: $p_1+p_2+p_3+\ldots+p_n+\ldots=1$. Для случайной величины, равной числу очков на верхней грани игральной кости, после её подбрасывания, таблица распределения имеет вид:

x_i:	1	2	3	4	5	6
p_i:	1/6	1/6	1/6	1/6	1/6	1/6

Функция распределения

Функцией распределения $F(x)$ вещественного аргумента x называется вероятность $F(x)=P(X<x)$. Построим $F(x)$ для случайной величины, равной числу очков на верхней грани игральной кости после её подбрасывания. При отрицательных аргументах x и даже при $x<1$ значение $F(x)=0$, потому что случайная величина никак не сможет принять значение меньшее x. При $x=1$, также $F(x)=0$, поскольку в определении функции распределения использовано строгое неравенство (событие $X<1$ невозможное). Если $1<x\le 2$, то $F(x)=1/6$, потому что событие $X<x$ при $1<x\le 2$ из-за дискретности возможных значений случайной величины совпадает с событием $X=1$, вероятность которого равна 1/6. Если $2<x\le 3$, то $F(x)=1/6+1/6=1/3$, потому что событие $X<x$ при $2<x\le 3$ совпадает с событием, заключающемся в наступлении хотя бы одного из двух элементарных событий: $X=1$ или $X=2$, а по аксиоме сложения вероятность такого события равна сумме вероятностей этих элементарных событий. Если $3<x\le 4$, то $F(x)=1/6+1/6+1/6=1/2$, потому что событие $X<x$ при $3<x\le 4$ совпадает с событием, заключающемся в наступлении хотя бы одного из трёх элементарных событий: $X=1$, $X=2$ или $X=3$. И так далее. Эта функция кусочно-заданная и её можно записать с помощью условного оператора (этот термин из программирования):

$$F(x)=\begin{cases} 0, \text{если } x \le 1, \\ 1/6, \text{если } 1 < x \le 2, \\ 1/3, \text{если } 2 < x \le 3, \\ 1/2, \text{если } 3 < x \le 4, \\ 2/3, \text{если } 4 < x \le 5, \\ 5/6, \text{если } 5 < x \le 6, \\ 1, \text{если } x > 6. \end{cases}$$

То, что при $x>6$ функция распределения $F(x)=1$, объясняется тем, что в этом случае событие $X<x$ является достоверным. На рисунке 10 изображён график этой функции.

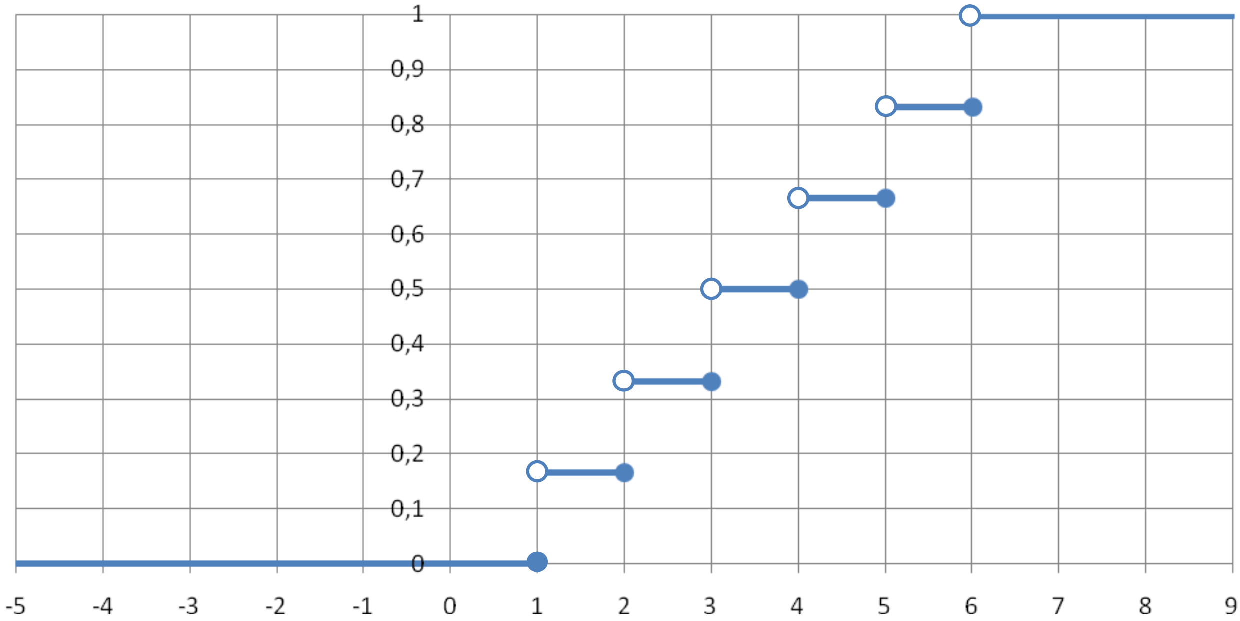

Рис.10. График функции распределения случайной величины, порождённой бросанием игральной кости.

Свойства функции распределения

1. Ограниченность: $0 \leq F(x) \leq 1$. Действительно это так, потому что $F(x)$ — это вероятность.
2. Монотонность: функция распределения $F(x)$ не убывает. Действительно, если $a<b$, то из выполнения события $X<a$ будет следовать выполнение события $X<b$, и по монотонности вероятности $P(X<a) \leq P(X<b)$, что равносильно неравенству $F(a) \leq F(b)$.
3. Если $a<b$, то $P(a \leq X<b)=F(b)-F(a)$. Действительно, событие $X<b$ может быть представлено в виде объединения двух несовместных событий $(X<b)=(X<a) \cup (a \leq X<b)$, что по аксиоме сложения вероятностей приводит к равенству: $P(X<b)=P(X<a)+P(a \leq X<b) \Leftrightarrow F(b)=F(a)+P(a \leq X<b) \Leftrightarrow P(a \leq X<b)=F(b)-F(a)$.
4. $\lim\limits_{x \to -\infty} F(x)=0$. Действительно $\lim\limits_{x \to -\infty} F(x)=P(X<-\infty)$, что невозможно.
5. $\lim\limits_{x \to +\infty} F(x)=1$. Действительно $\lim\limits_{x \to +\infty} F(x)=P(X<+\infty)$, что верно для любого X.
6. Функция $F(x)$ непрерывна слева. Это свойство следует из того, что для любого числа a (даже, если a — возможное значение дискретной случайной величины), при достаточно малом $\varepsilon>0$ (таком, чтобы промежуток $(a-\varepsilon;a)$ не содержал других возможных значений этой дискретной случайной величины) события $X<a$ и $X<a-\varepsilon$ неразличимы.
7. Функция $F(x)$ имеет разрывы первого рода (скачки) в возможных значениях дискретных случайных величин. Причем величины этих скачков равны вероятностям, с которыми реализуются соответствующие возможные значения случайной величины. Действительно, если $x=x_i$ возможное значение случайной величины X, то для любого достаточно малого числа $\varepsilon>0$, такого, что промежуток $(x_i;\ x_i+\varepsilon)$ не будет содержать других возможных значений случайной величины, событие $X<x_i+\varepsilon$ может быть представлено в виде объединения трёх несовместных событий $(X<x_i+\varepsilon)=(X<x_i) \cup (X=x_i) \cup (x_i<X<x_i+\varepsilon)$, что по аксиоме сложения вероятностей приводит к равенству: $P(X<x_i+\varepsilon)=P(X<x_i)+P(X=x_i)+P(x_i<X<x_i+\varepsilon) \Leftrightarrow$ $\Leftrightarrow F(x_i+\varepsilon)=F(x_i)+p_i+0 \Leftrightarrow F(x_i+\varepsilon)-F(x_i)=p_i$. В силу этого свойства по функции

распределения дискретной случайной величины можно однозначно восстановить её ряд распределения.

Вырожденная случайная величина

Вырожденной называется такая случайная величина, которая всегда принимает одно и то же значение. Другими словами, та, которая не случайная. Пусть это значение равно c. Эту «случайную» величину легко получить, подбрасывая кубик, у которого на всех гранях написано одно и то же число: c. В виде формулы определение выглядит так: $P(X=c)=1$. Её таблица распределения очень короткая:

x_i:	c
p_i:	1

Функция распределения такой величины имеет вид: $F(x)=\begin{cases}0, \text{если } x \le c,\\ 1, \text{если } x > c,\end{cases}$ так как событие $X<c$ невозможное, а $X \ge c$ достоверное. График функции распределения вырожденной случайной величины изображён на рисунке 11.

Рис. 11. График функции распределения вырожденной случайной величины.

Распределение Бернулли

Распределение Бернулли (с параметром p) имеет случайная величина, равная количеству гербов выпадающих при однократном бросании несимметричной монеты. Это количество гербов может быть равно либо 1, если герб выпал в этом единственном подбрасывании, либо 0, если выпала решка. Если вероятность выпадения герба равна p (и, соответственно, решки: $1-p$), то таблица распределения будет иметь вид:

x_i:	0	1
p_i:	$1-p$	p

Такую случайную величину ещё называют *индикатором* события, поскольку она показывает произошло событие или нет. Её функция распределения: $F(x)=\begin{cases}0, \text{если } x \le 0,\\ 1-p, \text{если } 0<x\le 1,\\ 1, \text{если } x>1.\end{cases}$ График этой функции показан на рисунке 12.

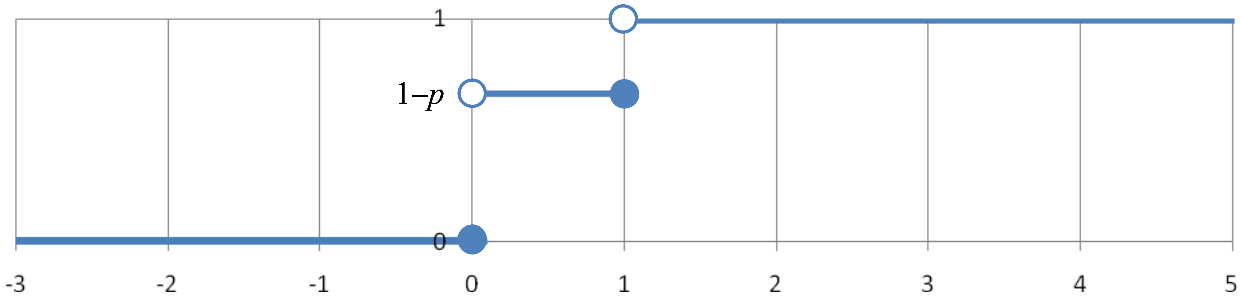

Рис. 12. График функции распределения случайной величины, имеющей распределение Бернулли.

Биномиальное распределение

Биномиальное распределение (с параметрами n и p) имеет случайная величина, равная количеству гербов выпадающих при многократных независимых бросаний несимметричной монеты. Если вероятность выпадения герба в одном отдельно взятом испытании равна p (решки: $1-p$) и число бросаний равно n, то по формуле Бернулли вероятность того, что в n бросаниях выпадут ровно m гербов равно $P_n(m) = C_n^m p^m (1-p)^{n-m}$. В n бросаниях число гербов может выпасть от 0 до n раз. Таблица распределения такой дискретной случайной величины имеет вид:

x_i:	0	1	2	…	m	…	$n-1$	n
p_i:	$(1-p)^n$	$n(1-p)^{n-1}p$	$\frac{n(n-1)}{2}(1-p)^{n-2}p^2$	…	$C_n^m(1-p)^{n-m}p^m$	…	$n(1-p)p^{n-1}$	p^n

Распределение получило такое название в связи с тем, что вероятности, записанные в нижней строке ряда распределения, являются слагаемыми в разложении бинома Ньютона:

$$1 = 1^n = ((1-p)+p)^n = \sum_{m=0}^{n} C_n^m (1-p)^{n-m} p^m =$$

$$= (1-p)^n + n(1-p)^{n-1}p + \frac{n(n-1)}{2}(1-p)^{n-2}p^2 + \ldots + C_n^m(1-p)^{n-m}p^m + \ldots + n(1-p)p^{n-1} + p^n.$$

Это дополнительно доказывает, что для данного распределения сумма $\sum_{i=0}^{n} p_i = \sum_{m=0}^{n} C_n^m (1-p)^{n-m} p^m = 1$.

Построим функцию распределения для простейшего случая: $n=2$, $p=1/2$ (два бросания симметричной монеты). Тогда таблица распределения случайной величины будет таким:

x_i:	0	1	2
p_i:	1/4	1/2	1/4

Справедливости ради заметим, что такая короткая таблица распределения может быть поострена и без формулы Бернулли. Вероятность события $X=0$ (выпали обе решки) может быть вычислена по правилу произведения: $P(X=0)=(1-p)(1-p)=(1/2)\cdot(1/2)=1/4$. Аналогично считается $P(X=2)=p\cdot p=(1/2)\cdot(1/2)=1/4$. А событие $X=1$ противоположно объединению непересекающихся событий $(X=0)\cup(X=2)$ и его вероятность равна $P(X=1)=1-P((X=0)\cup(X=2))=1-(P(X=0)+P(X=2))=1-((1/4)+(1/4))=1-1/2=1/2$. Функция распределения окажется равной:

$$F(x) = \begin{cases} 0, \text{если } x \le 0, \\ 1/4, \text{если } 0 < x \le 1, \\ 3/4, \text{если } 1 < x \le 2, \\ 1, \text{если } x > 2. \end{cases}$$

Её график показан на рисунке 13. Если построить функцию распределения не точно по формуле Бернулли, а приближённо с использованием интегральной теоремы Муавра–Лапласа, то она будет равна:

$$F(x) = P(X < x) = P(-\infty \le X < x) \approx \Phi\left(\frac{x-np}{\sqrt{npq}}\right) - \lim_{a\to-\infty}\Phi\left(\frac{a-np}{\sqrt{npq}}\right) = \Phi\left(\frac{x-np}{\sqrt{npq}}\right) - 0 = \Phi\left(\frac{x-np}{\sqrt{npq}}\right).$$

Здесь, как и раньше $q=1-p$. Для симметричной монеты при $p=q=1/2$ приближённо функция распределения: $F(x) \approx \Phi\left(\frac{2x}{\sqrt{n}} - \sqrt{n}\right)$. Конечно, странно использовать приближённые формулы при малых значениях n, но посмотрите, как пунктирная линия, являющаяся графиком функции $\Phi\left(\sqrt{2}(x-1)\right)$, недалеко проходит от графика точно построенной функции распределения.

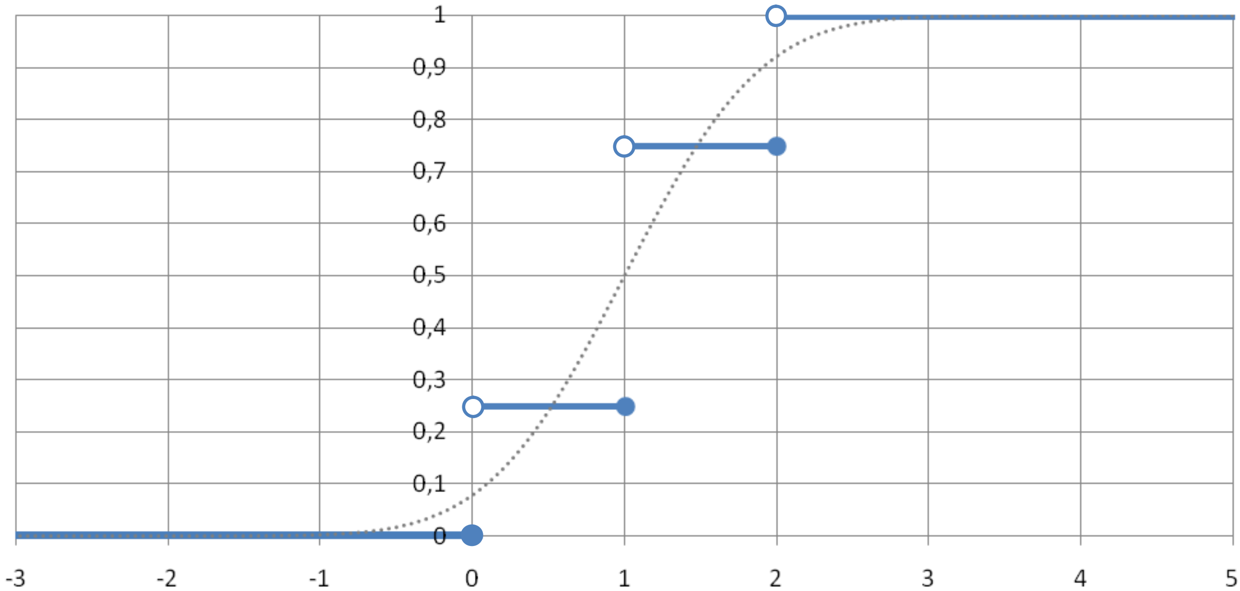

Рис. 13. Графики точной и приближённой функций распределения случайной величины, имеющей биномиальное распределение при n=2, p=1/2.

При n=3 и p=1/2 таблица распределения будет:

x_i:	0	1	2	3
p_i:	1/8	3/8	3/8	1/8

При p=1/2 в силу симметрии таблица распределения для n=3 тоже может быть поострена без формулы Бернулли. Вероятность события X=0 (выпали все три решки): $P(X=0)=(1/2)\cdot(1/2)\cdot(1/2)=1/8$. Вероятность выпадения всех трёх гербов: $P(X=3)=p\cdot p\cdot p=(1/2)\cdot(1/2)\cdot(1/2)=1/8$. А объединение событие $(X=1)\cup(X=2)$ противоположно объединению непересекающихся событий $(X=0)\cup(X=3)$ и его вероятность равна $P((X=1)\cup(X=2))=1-P((X=0)\cup(X=3))=1-(P(X=0)+P(X=3))=1-((1/8)+(1/8))=1-1/4=3/4$. Для симметричной монеты непересекающиеся события X=1 и X=2 равновероятны, поэтому $3/4=P((X=1)\cup(X=2))=P(X=1)+P(X=2)=2\cdot P(X=1)\Leftrightarrow P(X=1)=P(X=2)=3/8$. Функция распределения окажется равной: $F(x)=\begin{cases} 0, \text{если } x \le 0, \\ 1/8, \text{если } 0 < x \le 1, \\ 1/2, \text{если } 1 < x \le 2, \\ 7/8, \text{если } 2 < x \le 3, \\ 1, \text{если } x > 3. \end{cases}$ Приближённое выражение функции распределения: $F(x) \approx \Phi\left(\frac{2x}{\sqrt{3}} - \sqrt{3}\right)$. А графики показаны на рисунке 14. Естественно пунктирная линия отражает график точно вычисленной функции распределения $F(x)$ лучше, чем при n=2.

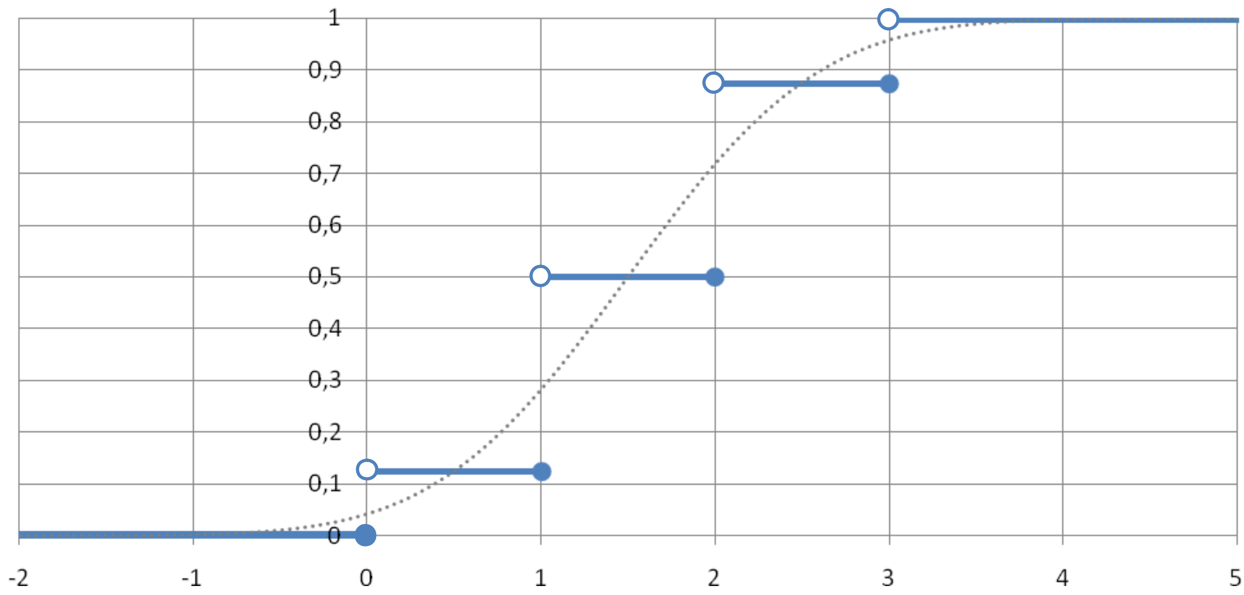

Рис. 14. Графики точной и приближённой функций распределения случайной величины, имеющей биномиальное распределение при n=3, p=1/2.

При n=4 даже в простейшем случае p=1/2 никакими эвристическими методами не удаётся построить таблицу распределения полностью. Здесь действительно нужно пользоваться формулой Бернулли или повторить её вывод для частного случая. Таблица распределения будет:

x_i:	0	1	2	3	4
p_i:	1/16	1/4	3/8	1/4	1/16

Функция распределения: $F(x)=\begin{cases} 0, \text{если } x \le 0, \\ 1/16, \text{если } 0 < x \le 1, \\ 5/16, \text{если } 1 < x \le 2, \\ 11/16, \text{если } 2 < x \le 3, \\ 15/16, \text{если } 3 < x \le 4, \\ 1, \text{если } x > 4. \end{cases}$ Приближённое выражение функции распределения: $F(x) \approx \Phi\left(x-2\right)$. Графики показаны на рисунке 15. Пунктирная линия ещё лучше, отражает график точно вычисленной функции распределения $F(x)$.

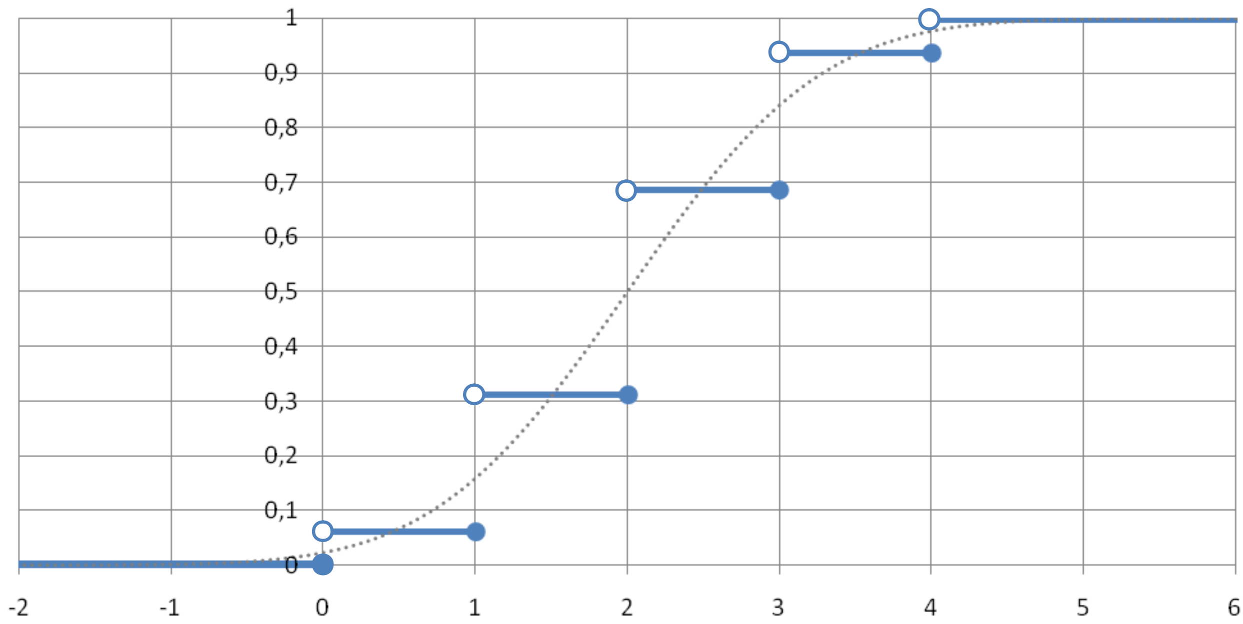

Рис. 15. Графики точной и приближённой функций распределения случайной величины, имеющей биномиальное распределение при n=4, p=1/2.

Геометрическое распределение

Геометрическое распределение (с параметром p) имеет случайная величина, равная количеству решек выпавших до первого выпадения герба при многократных независимых бросаниях несимметричной монеты. Если вероятность выпадения герба в одном отдельно взятом испытании равна p (решки: $1-p$), то в силу независимости испытаний, по формуле произведения вероятность того, что до первого герба выпадет ровно m решек равна: $P(X=m)=p\cdot(1-p)^m$. Такая дискретная случайная величина интересна тем, что у неё бесконечное количество возможных значений. Таблица распределения такой дискретной случайной величины имеет вид:

x_i:	0	1	2	3	…	m	…
p_i:	p	$p\cdot(1-p)$	$p\cdot(1-p)^2$	$p\cdot(1-p)^3$	…	$p\cdot(1-p)^m$	…

Распределение получило такое название в связи с тем, что вероятности, записанные в нижней строке ряда распределения, являются членами геометрической прогрессии. Её знаменателем является число $(1-p)\in(0;1)$ и такая прогрессия будет бесконечно убывающей. Для суммы членов такой прогрессии известно, как вычислить сумму бесконечного количества её членов: $S=\dfrac{b_1}{1-q}$. Здесь первый член прогрессии $b_1=p$, а её знаменатель $q=1-p$. Поэтому:

$\sum\limits_{i=0}^{\infty} p_i = \sum\limits_{m=0}^{\infty}\left(p\cdot(1-p)^m\right) = \dfrac{p}{1-(1-p)} = \dfrac{p}{p} = 1$. Выполнено условие нормированности, что косвенно подтверждает правильность построения таблицы распределения.

На рисунке 16 показан график функции распределения для простейшего случая: p=1/2 (монета симметричная).

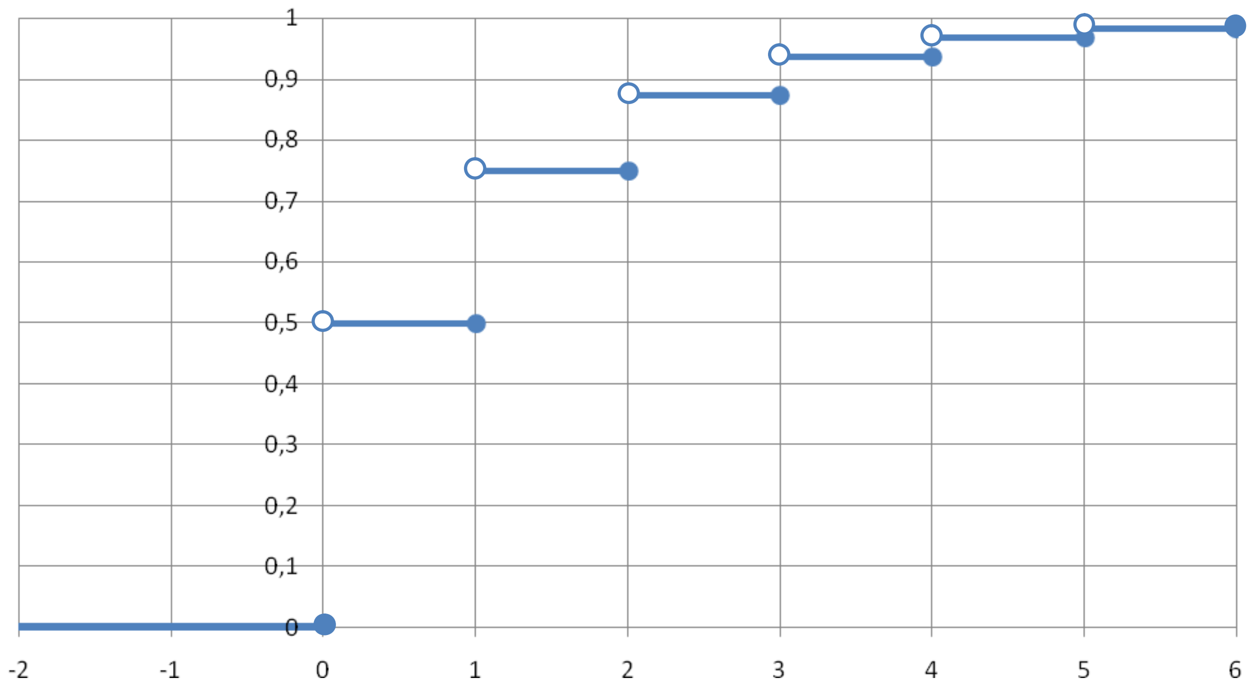

Рис. 16. График функции распределения случайной величины,
имеющей геометрическое распределение (при p=1/2).

Равномерно распределённая случайная величина

Вспомним задачу, которая привела к геометрическому способу вычисления вероятностей. Пусть точка x бросается на промежуток $[a;b)$, причём вероятности попадания этой точки в любые промежутки равной длины, являющиеся его частями, одинаковые. В таком примере случайной величиной X можно считать число x, отождествляя вещественное число и точку на числовой прямой. Назовём такую случайную величину равномерно распределённой на интервале $[a;b)$. Было показано, что при $a \le c \le d \le b$ вероятность $P(c \le x < d) = \dfrac{d-c}{b-a}$. Найдём функцию распределения такой случайной величины. При $x<a$ событие $X<x$ невозможное и $F(x)=P(X<x)=0$. При $x \ge b$ событие $X<x$ достоверное и $F(x)=P(X<x)=1$. При $x \in [a;b)$ событие $X<x$ совпадает с событием $a \le X < x$ и $F(x) = P(a \le X < x) = \dfrac{x-a}{b-a}$. С помощью условного оператора функцию распределения можно записать так: $F(x) = \begin{cases} 0, \text{если } x \le a, \\ \dfrac{x-a}{b-a}, \text{если } a < x < b, \\ 1, \text{если } x \ge b. \end{cases}$ График такой функции изображён на рисунке 17.

Непрерывные случайные величины

Если функция распределения случайной величины непрерывна, то такая случайная величина называется *непрерывной*. В частности, непрерывной будет случайная величина, равномерно распределённая на промежутке. Если на некотором промежутке функция распределения непрерывной случайной величины строго возрастает, то этот промежуток заполняется возможными значениями случайной величины сплошь. Для непрерывных случайных величин справедлив уже изученный ранее *парадокс нулевой вероятности*, согласно которому вероятность того, что непрерывно распределённая случайная величина

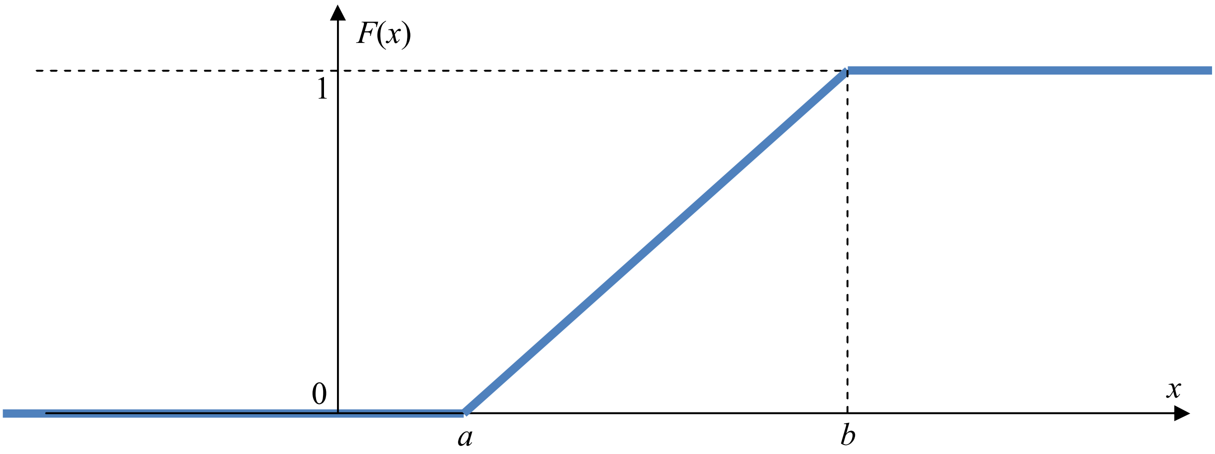

Рис. 17. График функции распределения случайной величины, равномерно распределённой на промежутке $[a;b)$.

примет конкретное значение равна нулю, то есть: $P(X=c)=0$. Докажем это. По аксиоме неотрицательности $P(X=c)\geq 0$. Кроме этого, по свойству функции распределения, для любого $\varepsilon>0$ верно ограничение $P(X=c)\leq P(c\leq X<c+\varepsilon)=F(c+\varepsilon)-F(c)$. Поэтому справедливо двойное неравенство: $0\leq P(X=c)\leq F(c+\varepsilon)-F(c)$. Перейдём в этом неравенстве к пределу: $\lim\limits_{\varepsilon\to+0} 0 \leq \lim\limits_{\varepsilon\to+0} P(X=c) \leq \lim\limits_{\varepsilon\to+0}\left(F(c+\varepsilon)-F(c)\right)$. Поскольку в этой записи констанами являются все числа кроме ε, а предел константы равен самой константе, то $0 \leq P(X=c) \leq \left(\lim\limits_{\varepsilon\to+0} F(c+\varepsilon)\right) - F(c)$. Так как функция $F(x)$ непрерывная, то по определению непрерывной функции её предел совпадает с её значением в предельной точке, поэтому: $0\leq P(X=c)\leq F(c)-F(c)\Leftrightarrow 0\leq P(X=c)\leq 0\Leftrightarrow P(X=c)=0$. В силу этого парадокса для равномерно распределённой на промежутке $[a;b)$ случайной величины не существенно, какие скобки использованы при записи промежутка: квадратные или круглые. Также для непрерывно распределённых случайных величин неважно, какие именно использовать знаки неравенств, при обозначении событий: строгие или не строгие. То есть $P(X\leq x)=P((X<x)\cup(X=x))=P(X<x)+P(X=x)=P(X<x)+0=P(X<x)=F(x)$. Аналогичное свойство верно и для двойных неравенств:

Не только $P(a\leq X<b)=F(b)-F(a)$, но и

$P(a\leq X\leq b)=P((a\leq X<b)\cup(X=b))=P(a\leq X<b)+P(X=b)=F(b)-F(a)+0=F(b)-F(a)$.

Поскольку $P(a\leq X<b)=P((X=a)\cup(a<X<b))=P(X=a)+P(a<X<b)=0+P(a<X<b)=P(a<X<b)$,

то $P(a<X<b)=P(a\leq X<b)=F(b)-F(a)$. Также

$P(a<X\leq b)=P((a<X<b)\cup(X=b))=P(a<X<b)+P(X=b)=F(b)-F(a)+0=F(b)-F(a)$.

Плотностью вероятности $f(x)$ называется производная функции распределения: $f(x)=F'(x)$. Поскольку $\lim\limits_{x\to-\infty} F(x)=0$ то $F(x)=\int\limits_{-\infty}^{x} f(t)dt$. Из свойства $\lim\limits_{x\to+\infty} F(x)=1$ следует условие нормированности для плотности вероятности: $\int\limits_{-\infty}^{+\infty} f(x)dx=1$. Поскольку функция распределения не убывающая, то её производная неотрицательная. Поэтому плотность вероятности тоже неотрицательная: $f(x)=F'(x)\geq 0$.

Выразим вероятность $P(a\leq X<b)$ через плотность вероятности $f(x)$ используя свойство аддитивности интеграла:

$$P(a\leq X<b)=F(b)-F(a)=\int\limits_{-\infty}^{b} f(x)dx-\int\limits_{-\infty}^{a} f(x)dx=\int\limits_{-\infty}^{a} f(x)dx+\int\limits_{a}^{b} f(x)dx-\int\limits_{-\infty}^{a} f(x)dx=\int\limits_{a}^{b} f(x)dx.$$

Также для непрерывных случайных величин $P(a < X < b) = P(a \le X \le b) = P(a < X \le b) = P(a \le X < b) = \int\limits_a^b f(x)dx$.

Из-за этого свойства функцию $f(x)$ и называют «плотность». Оно напоминает формулу, по которой масса тела равна интегралу по его объёму от его плотности.

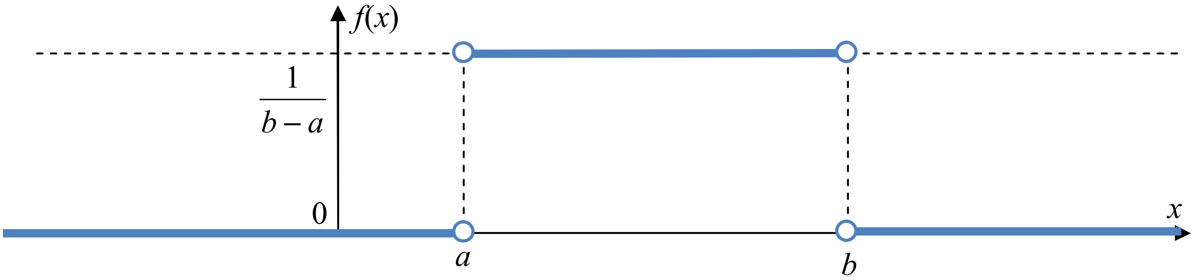

Рис. 18. График плотности вероятности случайной величины, равномерно распределённой на промежутке [a;b).

Найдем плотность вероятности равномерно распределённой на промежутке [a;b) случайной величины. Заметим, что в связи с парадоксом нулевой вероятности не столь важно входят концы a и b в промежуток или нет. Функция распределения равномерного распределения недифференцируема в точках излома графика: при $x=a$ и $x=b$. И поэтому производная функции распределения в этих двух точках неопределена: $f(x) = F'(x) = \begin{cases} 0, \text{если } x < a, \\ \dfrac{1}{b-a}, \text{если } a < x < b, \\ 0, \text{если } x > b. \end{cases}$ График плотности вероятности равномерного распределения показан на рисунке 18. Площадь фигуры под таким графиком равна площади прямоугольника: $(b-a)\cdot\dfrac{1}{b-a} = 1$, что подтверждает выполнение условия нормированности.

Нормально распределённая случайная величина

Нормально распределённой случайной величиной (с параметрами a и σ) называется такая случайная величина, у которой функция распределения равна $F(x) = \Phi\left(\dfrac{x-a}{\sigma}\right)$, где $\sigma>0$.

Плотность вероятности такой случайной величины, согласно правилу дифференцирования сложной функции, равна:

$$f(x) = F'(x) = \left(\Phi\left(\frac{x-a}{\sigma}\right)\right)' = \Phi'\left(\frac{x-a}{\sigma}\right)\cdot\left(\frac{x-a}{\sigma}\right)' = \varphi\left(\frac{x-a}{\sigma}\right)\cdot\frac{1}{\sigma} = \frac{1}{\sigma\sqrt{2\pi}}\cdot\exp\left(-\frac{(x-a)^2}{2\sigma^2}\right).$$

Качественно графики плотности вероятности и функции распределения выглядят так же как и графики $\varphi(x)$ и $\Phi(x)$, соответственно. График плотности вероятности $f(x)$ нормального распределения симметричен относительно вертикальной прямой $x=a$. Функция $f(x)$ положительная, имеет при $x=a$ максимум, совпадающий с наибольшим значением. Оно равно $y_{\max} = \dfrac{1}{\sigma}\cdot\varphi(0) = \dfrac{1}{\sigma\sqrt{2\pi}}$. График имеет две точки перегиба при $x=a-\sigma$ и $x=a+\sigma$. Значение функции в точках перегиба равно $y_{перегиба} = \dfrac{\varphi(-1)}{\sigma} = \dfrac{\varphi(1)}{\sigma} = \dfrac{1}{\sigma\sqrt{2\pi e}}$. Горизонтальная ось является асимптотой графика функции $\varphi(x)$. Покажем выполнение условия нормированности, для чего сделаем в несобственном интеграле $\int_{-\infty}^{+\infty}\varphi\left(\frac{x-a}{\sigma}\right)\cdot\frac{1}{\sigma}dx$ замену переменной: $t=\frac{x-a}{\sigma}\Leftrightarrow x=a+\sigma\cdot t\Rightarrow dx=\sigma\cdot dt$, тогда при $x\to-\infty \Leftrightarrow t\to-\infty$, при $x\to+\infty \Leftrightarrow t\to+-\infty$. Тогда $\int_{-\infty}^{+\infty}\varphi\left(\frac{x-a}{\sigma}\right)\cdot\frac{1}{\sigma}dx=\int_{-\infty}^{+\infty}\varphi(t)\cdot\frac{1}{\sigma}\cdot\sigma\cdot dt=\int_{-\infty}^{+\infty}\varphi(t)dt=1$. График функции распределения $F(x)$ нормально распределённой случайной величины центрально симметричен относительно своей точки перегиба при $x=a$, в которой $F(a)=1/2$. Это с арифметической точки зрения выражается равенством $F(a+x)+F(a-x)=1$. График $F(x)$ имеет две горизонтальные асимптоты: $y=0$ при $x\to-\infty$ и $y=1$ при $x\to+\infty$.

На рисунках 19 и 20 показано сравнение графиков плотностей вероятностей и функций распределения двух нормально распределённых случайных величин. У одной величины a=6, σ=1 у другой a=6, σ=2. Видно, что увеличение параметра a сдвигает графики направо по горизонтали, а параметр σ отвечает за пологость кривых.

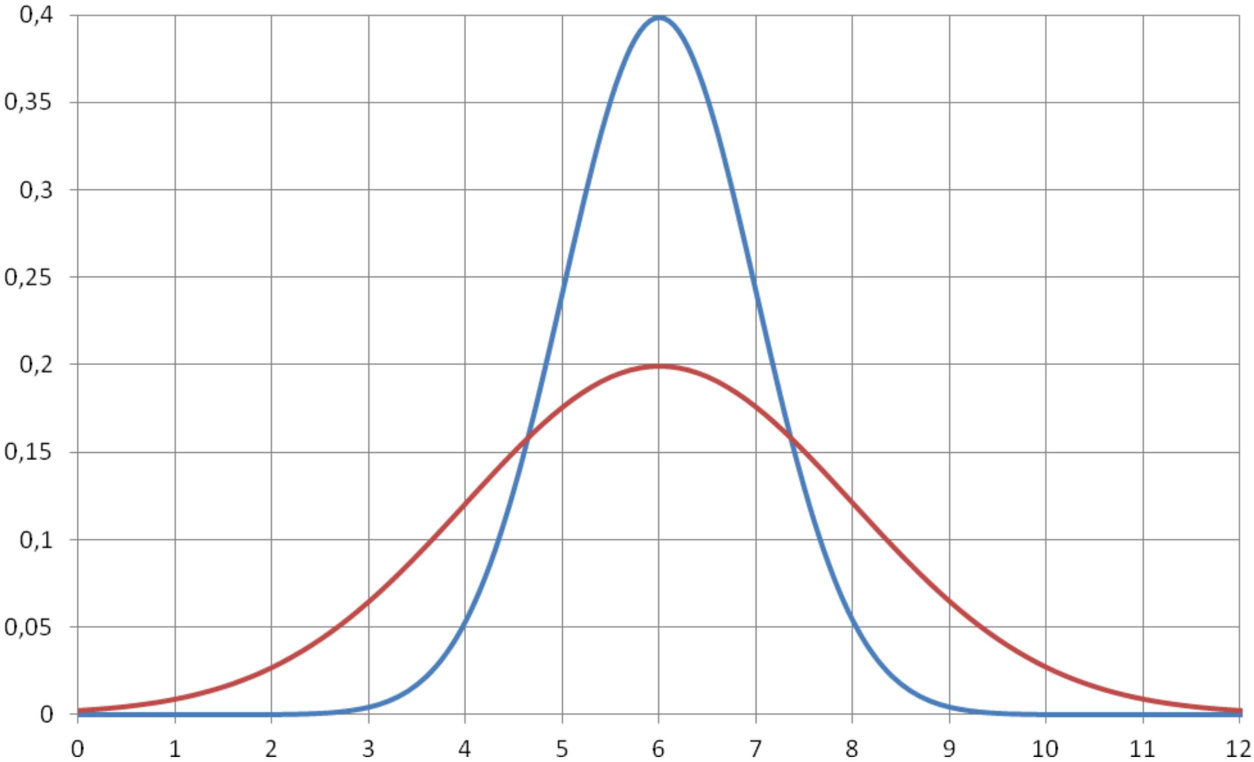

Рис. 19. Графики плотностей вероятностей двух нормально распределенных случайных величин при a=6, σ=1 и при a=6, σ=2.

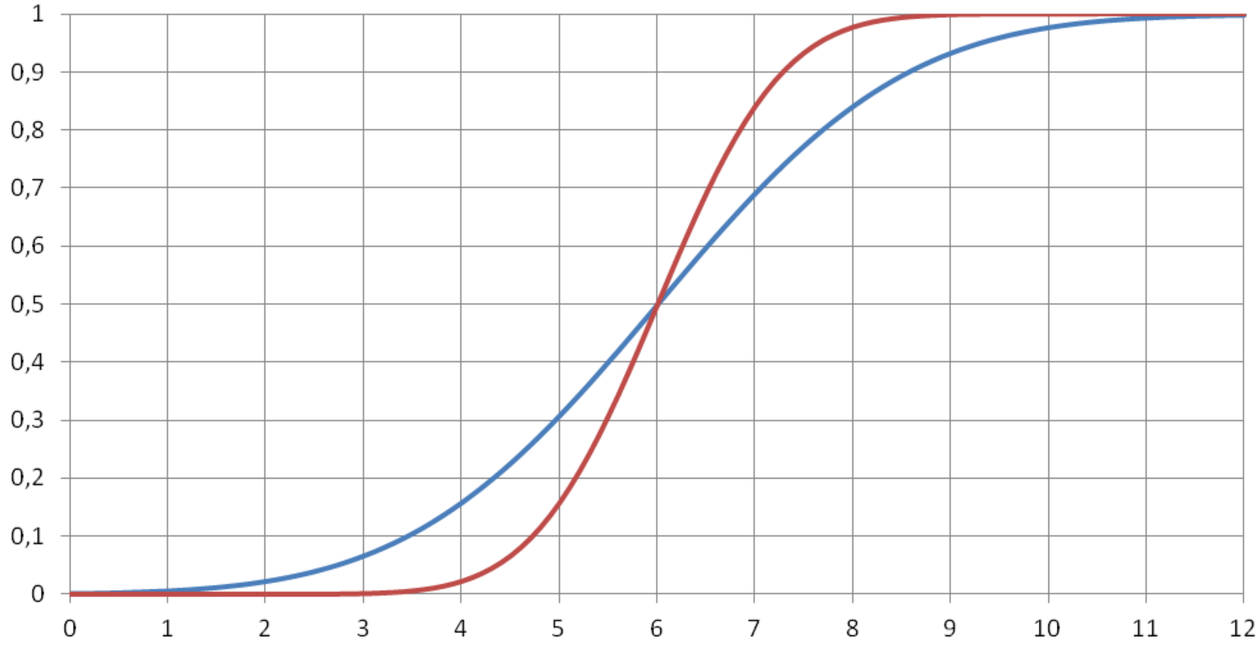

Рис. 20. Графики функций распределения двух нормально распределенных случайных величин при a=6, σ=1 и при a=6, σ=2.

Нормальное распределение интересно тем, что большинство случайных величин встречающихся на практике (и особенно в биологии) распределено по законам близким к нормальному (отсюда и его название). Графики плотностей их вероятностей, как и нормального распределения, тоже имеют вид колокола (возможно чуть-чуть не симметричного относительно вертикальной прямой, проходящей через вершину графика, см. рис. 21 и 22). Принципиальным отличием плотностей вероятностей этих случайных величин от функции $\varphi(x)$ является то, что их возможные значения заполняют некоторый ограниченный промежуток, а не всю числовую прямую. Например, рост человека не может быть отрицательным и никто никогда не видел людей выше трёх метров.

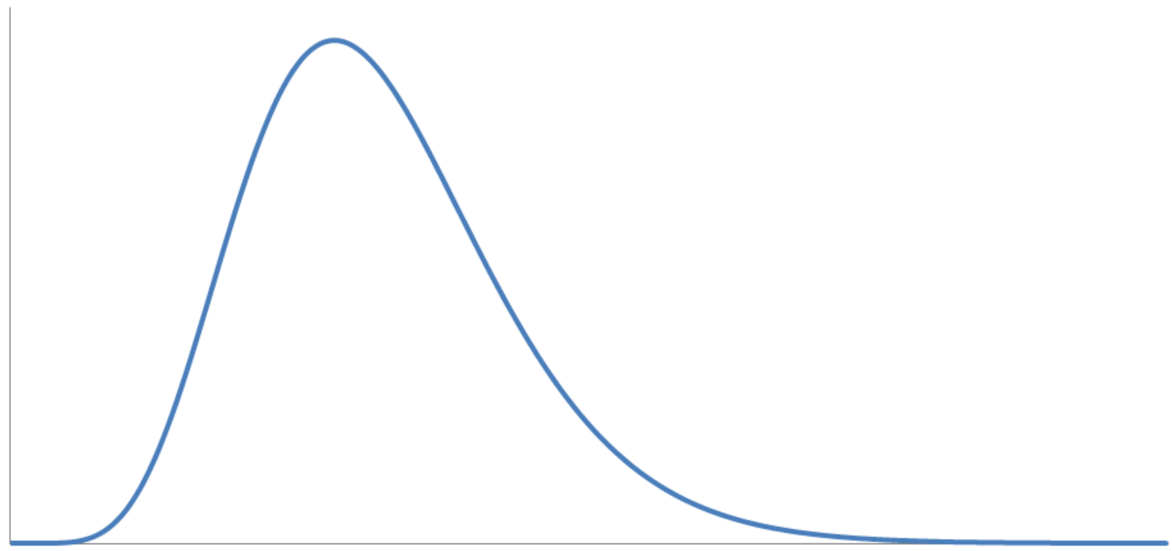

Рис. 21. Плотность вероятности случайной величины, близкой к нормальной.

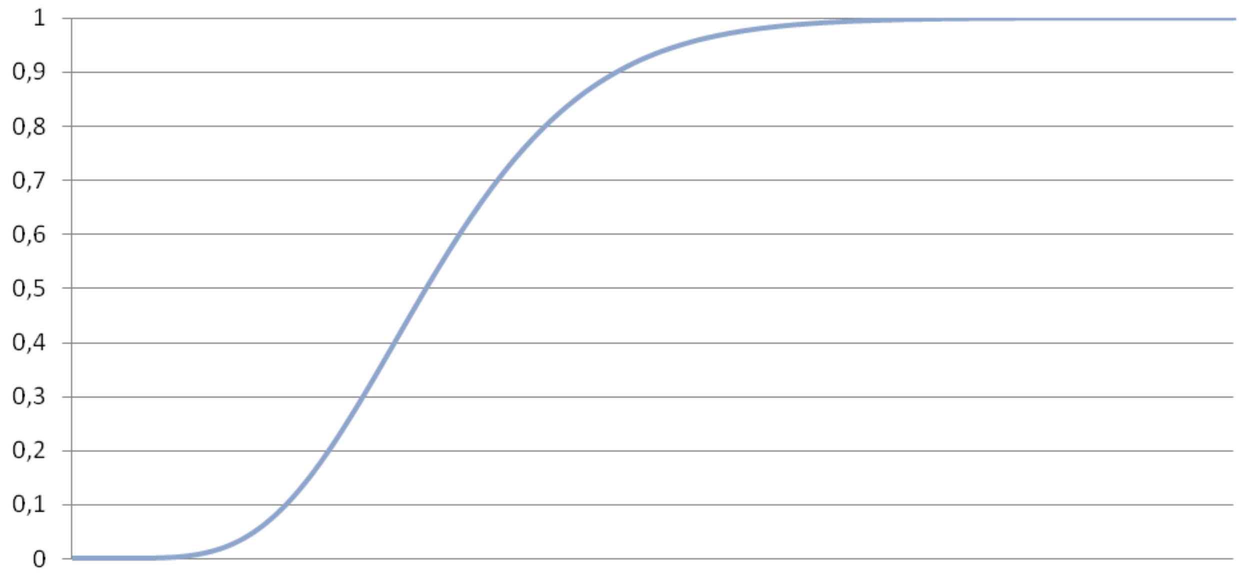

Рис. 22. Функция распределения случайной величины, близкой к нормальной.

Обратим ещё раз внимание на графики на рисунке 19. Параметр σ характеризует разбросанность нормально распределённой случайной величины относительно параметра a. Увеличение параметра σ в k раз не только уменьшает в k раз высоту максимума графика плотности вероятности, но ещё и сопровождается увеличением значений этой функции на «хвостах», то есть вдали от $x=a$. С геометрической точки зрения это понятно: по условию нормированности площадь под графиком плотности вероятности должна равняться единице, то есть не изменяться при уменьшении высоты максимума графика. Естественно, такое уменьшение должно сопровождаться увеличением значений плотности вероятности по краям (иначе площадь под графиком уменьшится). Эту особенность нормально распределённых случайных величин можно проиллюстрировать любопытным примером. Выше было сказано, что большинство случайных величин в биологии распределены по законам, близким к нормальному. К этим случайным величинам можно отнести и врождённые способности человека в какой-либо области. Оказывается, что разброс врождённых способностей у мужчин гораздо больше, чем у женщин. Это связано с тем, что на воспроизводство потомства женщина тратит в сотни раз больше времени, чем мужчина. А это, в свою очередь, приводит к тому, что желательно, чтобы в процессе воспроизводства участвовали все женщины. А для этого не должно быть женщин с низкими способностями. Поэтому рассеяние врождённых способностей женщин «делается» природой как можно меньше. В итоге, рассеяние способностей мужчин оказывается больше рассеяния способностей женщин. В жизни заметных успехов достигают люди со способностями, значительно превышающими средний уровень. И, как следует из того, что на правом «хвосте» график плотности вероятности мужских способностей оказывается выше, чем женских, такими людьми оказываются в подавляющем большинстве мужчины. В частности, подобное явление можно проследить в интеллектуальных видах спорта, где преимущество мужчин в физической силе не имеет принципиального значения, например, в шахматах. Шахматы удобны для иллюстрации тем, что в них принято вычислять силу игроков с помощью рейтинга Эло. Отсортировав шахматистов по убыванию рейтинга, можно увидеть, что в первой сотне лучших игроков мира всего одна женщина. Во второй сотне ещё одна. И только в третьей сотне лучших шахматистов мира более одной женщины. Вышесказанное не означает, что женщины играют в шахматы хуже мужчин. Если задать вопрос: «сможет ли женщина стать чемпионом мира по шахматам в тотале (то есть и среди мужчин тоже)?», то ответом будет «да, сможет, но в первой сотне лучших шахматистов мира она так и останется единственной женщиной». Заметим ещё что, график плотности вероятностей мужских способностей оказывается выше не только на правом, но и на левом «хвосте». То есть мужчин с очень низкими способностями тоже гораздо больше, чем женщин с очень низкими способностями.

Плотность вероятности дискретной случайной величины

Функция распределения $F(x)$ дискретной случайной величины на участках непрерывности равна константам, производные от которых равны нулю, а в точках разрыва недифференцируема. Такую производную функции распределения в качестве плотности вероятности рассматривать бессмысленно. Плотностью вероятности могла бы быть функция $f(x)$ такая, что: во-первых, она была бы действительно равна нулю в тех точках, где равна нулю $F'(x)$, а во-вторых, выполнялось бы равенство: $F(x) = \int_{-\infty}^{x} f(t)dt$. Разумеется, такой функции $f(x)$ не существует. Чтобы получить функцию, которую имеет смысл называть плотностью вероятности, можно использовать *обобщённую функцию*: дельта-функцию Дирака $\delta(x)$.

Дельта-функцией Дирака $\delta(x)$ называется такая обобщённая функция, которая равна нулю во всех точках кроме нуля, и при этом определённый интеграл по любому промежутку, начинающемуся с нуля, от этой функции равен единице. В виде формул это выглядит так: $\delta(x)=0$, если $x\neq0$ и $\int_{0}^{\varepsilon} \delta(x)dx = 1$, если $\varepsilon>0$. «График» такой обобщённой функции показан на рисунке 22. Фактически это точечный всплеск с бесконечной высотой, но при этом единичной интенсивности. Площадь фигуры под графиком остаётся равной единице при стремлении толщины прямоугольника под графиком к нулю, а высоты к бесконечности.

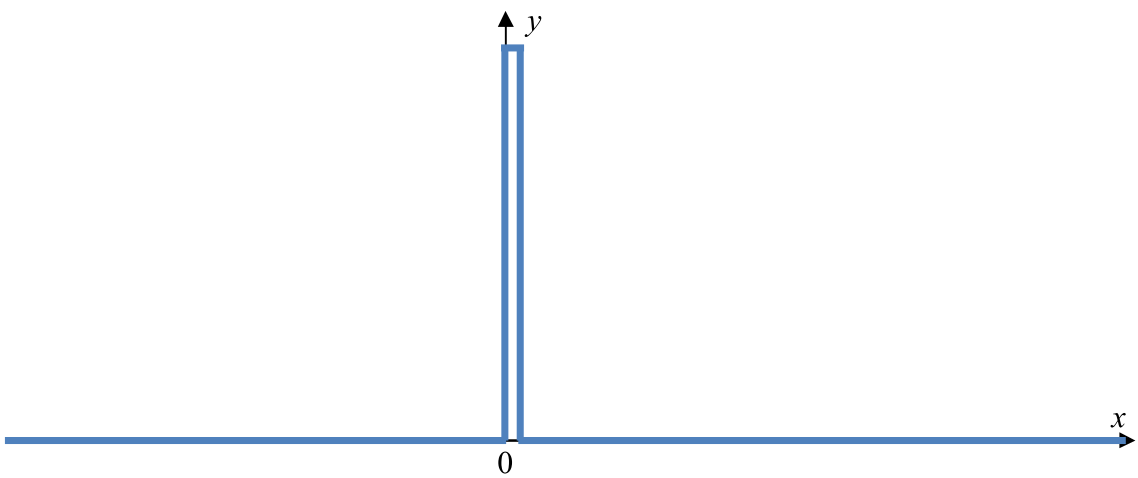

Рис. 22. Схематический график обобщённой дельта-функции Дирака.

Например, с помощью дельта-функции Дирака можно записать плотность вероятности вырожденной случайной величины: $f(x)=\delta(x-c)$. Действительно, при $x\le c$ функция распределения $F(x)=\int_{-\infty}^{x}\delta(t-c)dt=\int_{-\infty}^{x}0dt=0$. А при $x>c$ будет справедливо

$F(x)=\int_{-\infty}^{x}\delta(t-c)dt=\int_{-\infty}^{c}\delta(t-c)dt+\int_{c}^{x}\delta(t-c)dt$. Во втором интеграле сделаем замену переменной: $u=t-c\Rightarrow du=dt$; $t=c\Rightarrow u=0$; $t=x\Rightarrow u=x-c$. Тогда: $F(x)=\int_{-\infty}^{c}0dt+\int_{0}^{x-c}\delta(u)du=0+1=1$.

Чтобы получить плотность вероятности произвольной дискретной случайной величины X, для которой $P(X=x_i)=p_i$, необходимо использовать знак суммирования: $f(x)=\sum_i p_i\delta(x-x_i)$. Здесь число слагаемых совпадает с количеством возможных значений случайной величины X и может быть как конечным, так и бесконечным. Действительно, если x меньше либо равно наименьшего из возможных значений, то функция распределения $F(x)=\int_{-\infty}^{x}0dt=0$. А при переходе переменной x через возможное значение x_k функция распределения $F(x)$ увеличится на величину $F(x_k+\varepsilon)-F(x_k)=\int_{x_k}^{x_k+\varepsilon}f(x)dx=\int_{x_k}^{x_k+\varepsilon}\sum_i p_i\delta(x-x_i)dx=\sum_i p_i\int_{x_k}^{x_k+\varepsilon}\delta(x-x_i)dx=\sum_i p_i\delta_{ik}=$

$=p_1\cdot 0+p_2\cdot 0+p_3\cdot 0+\ldots+p_k\cdot 1+\ldots+p_n\cdot 0+\ldots=p_k$, где $0<\varepsilon<x_{k+1}-x_k$. В выкладках использовано обозначение $\delta_{ik}=\begin{cases}1, \text{если } i=k,\\ 0, \text{если } i\ne k.\end{cases}$ Символ δ_{ik} называется дельтой Кронекера.

Смешанные случайные величины

Если функция распределения случайной величины имеет хотя бы один разрыв, и её возможные значения заполняют хотя бы один промежуток сплошь, то она называется *смешанной* случайной величиной.

В качестве примера смешанной случайной величины приведём время ожидания разрешающего сигнала светофора на пешеходном переходе при подходе пешехода к перекрёстку в случайный момент. Допустим, на рассматриваемом перекрёстке зелёный и красный сигналы светофора включаются каждый на 30 секунд, а жёлтый (включающийся два раза за цикл) горит 5 секунд. По условию, человек подходит к перекрёстку в произвольный момент 70-секундного цикла. Если это произошло при включённом зелёном сигнале, то время ожидания начала перехода улицы будет равно нулю. Вероятность такого события равна $P(X=0)=30/70=3/7$. Если человек подошёл к перекрёстку в тот момент, когда включён один из запрещающих сигналов светофора (напомню, что начинать движение разрешается только по зелёному сигналу светофора!), то ему придётся подождать некоторое время от нуля до 40 секунд. Поскольку вероятность появления пешехода на перекрёстке в любые промежутки времени равной длины, одинаковая, то оставшиеся 1–3/7=4/7 вероятности равномерно распределятся на промежутке от 0 до 40 секунд. График функции распределения такой случайной величины показан на рисунке 23.

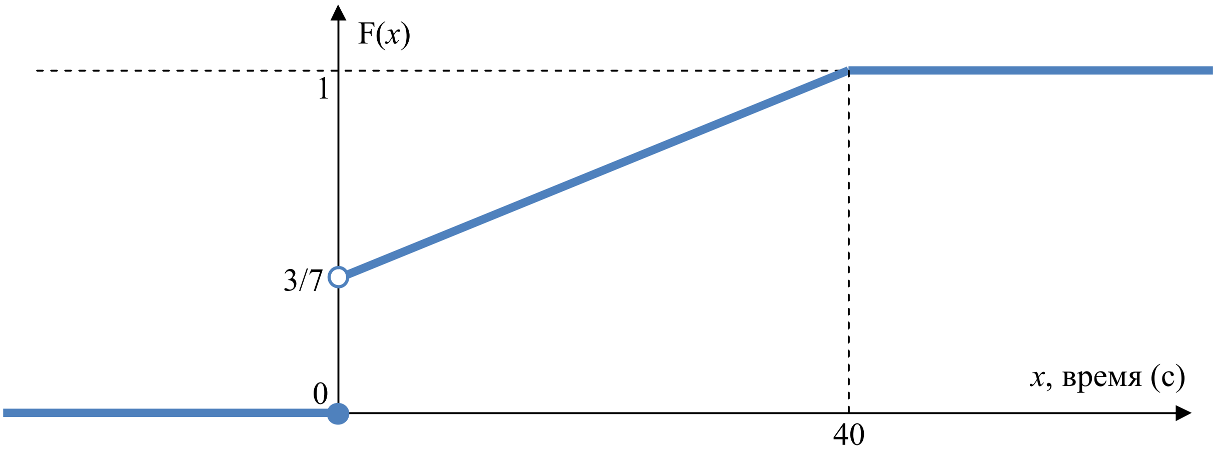

Рис. 23. График функции распределения смешанной случайной величины, равной времени ожидания перехода улицы.

Аналитически функцию распределения можно задать так: $F(x)=\begin{cases}0,\text{если } x\le 0,\\ \dfrac{30+x}{70},\text{если } 0<x\le 40,\\ 1,\text{если } x>40.\end{cases}$

Плотность вероятности этой смешанной случайной величины можно задать с помощью дельта-функции Дирака: $f(x)=\dfrac{3}{7}\cdot\delta(x)+\begin{cases}0,\text{если}\, x<0,\\ \dfrac{1}{70},\text{если}\, 0<x<40,\\ 0,\text{если } x>40.\end{cases}$ Её схематический график показан на рисунке 24.

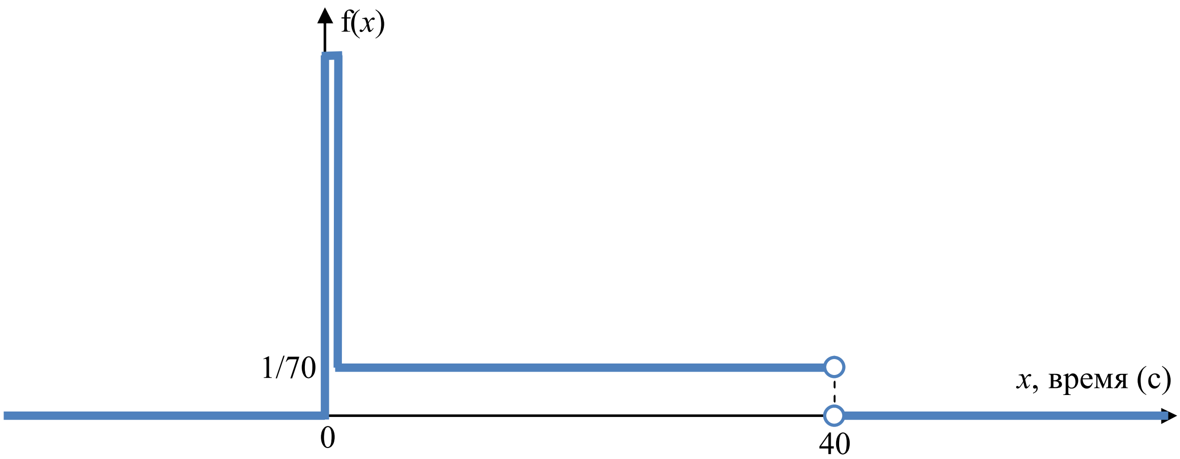

Рис. 24. Схематический график плотности вероятности смешанной случайной величины, равной времени ожидания перехода улицы.

Математическое ожидание

Математическим ожиданием случайной величины X с плотностью вероятности $f(x)$ называется несобственный интеграл: $EX=MX=\int\limits_{-\infty}^{+\infty} xf(x)dx$, если сходится несобственный интеграл $\int\limits_{-\infty}^{+\infty} |xf(x)|\,dx$. Исторически сложилось, что для математического ожидания существуют два обозначения EX и MX.

Почему формула такая? Математик принимает её за данное, как определение, а представители других наук видят в ней аналогию с формулой для нахождения координаты центра тяжести фигуры с плотностью $f(x)$ в точке с координатой x. В связи с этой аналогией возникает вопрос: «Неужели центр тяжести фигуры может не существовать?». Действительно, если несобственный интеграл $\int\limits_{-\infty}^{+\infty} |xf(x)|\,dx$ окажется расходящимся, то из определения следует, что у случайной величины нет математического ожидания. Разумеется, у ограниченных фигур центр тяжести есть всегда, этот интеграл может расходиться только, если возможные значения случайной величины неограниченны. В обычной жизни человек не встречается с неограниченными фигурами, а их свойства могут отличаться от свойств привычных фигур, и не стоит удивляться тому, что у неограниченных фигур может не быть центра тяжести.

Выведем формулу для математического ожидания произвольной дискретной случайной величины X, для которой $P(X=x_i)=p_i$:

$$EX=\int\limits_{-\infty}^{+\infty} xf(x)dx=\int\limits_{-\infty}^{+\infty} x\sum_i p_i\delta(x-x_i)dx=\sum_i p_i\int\limits_{-\infty}^{+\infty} x\delta(x-x_i)dx=\sum_i p_i\int\limits_{-\infty}^{+\infty} x_i\delta(x-x_i)dx=$$

$$=\sum_i p_i x_i\int\limits_{-\infty}^{+\infty}\delta(x-x_i)dx=\sum_i p_i x_i\cdot 1=\sum_i x_i p_i.$$ Как следует из определения, математическое ожидание дискретной случайной величины существует тогда, когда ряд $\sum_i |p_i x_i|$ будет сходящимся.

Примеры нахождения математических ожиданий

Найдём математическое ожидание для вырожденной случайной величины, принимающей значение c с вероятностью единица: $EX=x_1\cdot p_1=c\cdot 1=c$. Получилось единственное возможное значение этой величины. Вряд ли кто-то ожидал чего-то другого.

Найдём математическое ожидание случайной величины, порождённой бросанием игральной кости: $EX=x_1\cdot p_1+x_2\cdot p_2+x_3\cdot p_3+x_4\cdot p_4+x_5\cdot p_5+x_6\cdot p_6=$
$=1\cdot(1/6)+2\cdot(1/6)+3\cdot(1/6)+4\cdot(1/6)+5\cdot(1/6)+6\cdot(1/6)=(1+2+3+4+5+6)/6=21/6=7/2=3{,}5$.
Получилось три с половиной. Какой в этом смысл? Неужели кто-то ожидает, что на кубике выпадет три с половиной очка? Нет. Ожидается, что <u>в среднем</u> выпадет три с половиной. Под этим подразумевается, что если совершить много испытаний и вычислить среднее арифметическое результатов, то получится число не сильно отличающееся от трёх с половиной.

Действительно, если совершить большое число N испытаний дискретной случайной величины, то частота f_i появления возможного значения x_i, будет приблизительно равна вероятности p_i, и $EX=\sum_i x_i p_i\approx\sum_{i=1}^{N} x_i f_i$. Величину $\sum_{i=1}^{N} x_i f_i$, всегда являющуюся <u>конечной</u> суммой, и поэтому всегда существующее, называют *выборочное среднее*. В опыте она может быть вычислена и для непрерывно распределённых случайных величин.

Найдём математическое ожидание случайной величины, имеющей распределение Бернулли с параметром p: $EX=x_1\cdot p_1+x_2\cdot p_2=0\cdot(1-p)+1\cdot p=p$. Запомним этот результат.

Найдём математическое ожидание случайной величины, имеющей геометрическое распределение с параметром p (в силу неотрицательности возможных значений необходимости рассматривать ряд из модулей нет):

$$EX=\sum_i p_i x_i=\sum_{m=0}^{\infty} mp(1-p)^m=\sum_{m=1}^{\infty} mp(1-p)^m=p\sum_{m=1}^{\infty} m(1-p)^m=p\sum_{m=1}^{\infty} mq^m=p\sum_{m=1}^{\infty} mqq^{m-1}=$$

$$pq\sum_{m=1}^{\infty} mq^{m-1}=pq\sum_{m=1}^{\infty}\left(q^m\right)'_q=pq\left(\sum_{m=1}^{\infty} q^m\right)'_q=pq\left(q+q^2+q^3+\ldots\right)'_q=pq\left(\frac{q}{1-q}\right)'_q=$$

$$=pq\cdot\frac{1\cdot(1-q)-(-1)\cdot q}{(1-q)^2}=pq\frac{1-q+q}{(1-q)^2}=\frac{pq}{(1-q)^2}=\frac{p(1-p)}{p^2}=\frac{q}{p}=\frac{1-p}{p}=\frac{1}{p}-1.$$ В расчётах использованы: стандартное обозначение q=1–p, формула для производной степенной функции, возможность почленно дифференцировать равномерно сходящийся функциональный ряд, формула суммы бесконечно убывающей геометрической прогрессии, правило дифференцирования частного. Если p очень мало, то человек часто подбирает такое натуральное число n, что $p\approx 1/n$. Тогда $1/p\approx n$, $EX\approx n-1\Leftrightarrow EX+1\approx n$. Вспомним смысл случайной величины, имеющей геометрическое распределение. Это количество неудач до первого успеха. Если добавить единицу, то получится номер испытания, когда был первый успех. Наверно никого не удивляет, что если вероятность успеха в одном испытании равна $1/n$, то первый успех в среднем придёт в n-ом испытании.

Найдём математическое ожидание случайной величины, равномерно распределённой на промежутке $[a;b)$, используя свойство аддитивности интеграла: $EX=\int\limits_{-\infty}^{+\infty} xf(x)dx=$

$$=\int\limits_{-\infty}^{a} xf(x)dx+\int\limits_{a}^{b} xf(x)dx+\int\limits_{b}^{+\infty} xf(x)dx=\int\limits_{-\infty}^{a} x\cdot 0dx+\int\limits_{a}^{b} x\cdot\frac{1}{b-a}dx+\int\limits_{b}^{+\infty} x\cdot 0dx=$$

$$=\int\limits_{-\infty}^{a} 0dx+\frac{1}{b-a}\int\limits_{a}^{b} x\,dx+\int\limits_{b}^{+\infty} 0dx=0+\frac{1}{b-a}\cdot\left.\left(\frac{x^2}{2}\right)\right|_a^b+0=\frac{1}{b-a}\cdot\left(\frac{b^2}{2}-\frac{a^2}{2}\right)=\frac{b^2-a^2}{2(b-a)}=$$

$$=\frac{(b-a)(b+a)}{2(b-a)}=\frac{a+b}{2}.$$ Получилась середина промежутка, на котором равномерно распределена случайная величина. Из представлений о свойствах центра тяжести фигуры должно было получиться то же самое.

Найдём математическое ожидание нормально распределённой с параметрами a и σ случайной величины. Для начала установим, существует ли оно. Докажем сходимость интеграла $\int\limits_{-\infty}^{+\infty}\left|xf(x)\right|dx=\int\limits_{-\infty}^{+\infty}\left|x\cdot\frac{1}{\sigma}\cdot\varphi\left(\frac{x-a}{\sigma}\right)\right|dx=\int\limits_{-\infty}^{+\infty}|x|\cdot\frac{1}{\sigma}\cdot\varphi\left(\frac{x-a}{\sigma}\right)dx$. Сделаем замену переменной: $t=\frac{x-a}{\sigma}\Leftrightarrow x=a+\sigma\cdot t\Rightarrow dx=\sigma\cdot dt$, тогда при $x\to-\infty\Leftrightarrow t\to-\infty$, при $x\to+\infty\Leftrightarrow t\to+\infty$. Тогда

$$\int\limits_{-\infty}^{+\infty}\left|xf(x)\right|dx=\int\limits_{-\infty}^{+\infty}|a+\sigma\cdot t|\cdot\frac{1}{\sigma}\cdot\varphi(t)\cdot\sigma\cdot dt=\int\limits_{-\infty}^{+\infty}|a+\sigma\cdot t|\cdot\varphi(t)dt\le\int\limits_{-\infty}^{+\infty}\left(|a|+\sigma\cdot|t|\right)\varphi(t)dt=$$

$$=|a|\int\limits_{-\infty}^{+\infty}\varphi(t)dt+\sigma\cdot\int\limits_{-\infty}^{+\infty}|t|\,\varphi(t)dt=|a|\cdot 1+\sigma\cdot 2\int\limits_{0}^{+\infty}t\varphi(t)dt=|a|+2\sigma\int\limits_{0}^{+\infty}t\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{t^2}{2}\right)dt.$$ Здесь использовано условие нормированности: $\int\limits_{-\infty}^{+\infty}\varphi(t)dt=1$. В последнем интеграле сделаем ещё одну замену переменных $u=-\frac{t^2}{2}\Rightarrow du=-\frac{2t\,dt}{2}=-t\,dt$, тогда при t=0$\Leftrightarrow u$=0, при $t\to+\infty\Leftrightarrow u\to-\infty$. Тогда

$$\int_{-\infty}^{+\infty}|xf(x)|dx \le |a| + 2\sigma\int_0^{-\infty} -\frac{1}{\sqrt{2\pi}}e^u du = |a| - \frac{2\sigma}{\sqrt{2\pi}}\int_0^{-\infty} e^u du = |a| - \sigma\sqrt{\frac{2}{\pi}}\lim_{b\to-\infty}\int_0^b e^u du =$$

$= |a| - \sigma\sqrt{\frac{2}{\pi}}\lim_{b\to-\infty} e^u\Big|_0^b = |a| - \sigma\sqrt{\frac{2}{\pi}}\lim_{b\to-\infty}\left(e^b - e^0\right) = |a| - \sigma\sqrt{\frac{2}{\pi}}(0-1) = |a| + \sigma\sqrt{\frac{2}{\pi}}$. Доказали, что интеграл меньше некоторого числа, следовательно он сходится. Теперь имеет смысл искать и само математическое ожидание:

$EX = \int_{-\infty}^{+\infty} xf(x)dx = \int_{-\infty}^{+\infty} x\cdot\frac{1}{\sigma}\cdot\varphi\left(\frac{x-a}{\sigma}\right)dx$. Сделаем ту же замену переменной: $t = \frac{x-a}{\sigma}$, тогда при $x\to-\infty$ ⇔ t→–∞, при $x\to+\infty$ ⇔ t→+∞.

$$EX = \int_{-\infty}^{+\infty}(a+\sigma\cdot t)\cdot\frac{1}{\sigma}\cdot\varphi(t)\cdot\sigma\cdot dt = \int_{-\infty}^{+\infty}(a+\sigma\cdot t)\cdot\varphi(t)dt = a\int_{-\infty}^{+\infty}\varphi(t)dt + \sigma\int_{-\infty}^{+\infty}t\cdot\varphi(t)dt = a\cdot 1 + \sigma\cdot 0 = a.$$

Здесь один из интегралов равен единице по условию нормированности, а другой нулю, как сходящийся интеграл от нечётной функции по симметричному относительно нуля промежутку. Итак, результат: *EX*=*a*. Этот результат и ожидался, поскольку *x*=*a* точка симметрии графика плотности вероятности нормально распределённой случайной величины. Фактически вопрос был только в существовании центра тяжести этой неограниченной фигуры.

Найдём математическое ожидание смешаной случайной величины, равной времени ожидания перехода улицы. Пусть плотность вероятности соответствует графику, изображённому на рис. 24 и выражается фомулой:

$$f(x) = \frac{3}{7}\cdot\delta(x) + \begin{cases} 0, \text{если } x < 0, \\ \frac{1}{70}, \text{если } 0 < x < 40, \\ 0, \text{если } x > 40. \end{cases} \text{Тогда}$$

$$EX = \int_{-\infty}^{+\infty} xf(x)dx = \int_{-\infty}^{+\infty} x\left(\frac{3}{7}\cdot\delta(x) + \begin{cases} 0, \text{если } x < 0, \\ \frac{1}{70}, \text{если } 0 < x < 40, \\ 0, \text{если } x > 40 \end{cases}\right)dx =$$

$$= \int_{-\infty}^{+\infty}\left(\frac{3}{7}\cdot x\delta(x) + \begin{cases} 0, \text{если } x < 0, \\ \frac{x}{70}, \text{если } 0 < x < 40, \\ 0, \text{если } x > 40 \end{cases}\right)dx = \int_{-\infty}^{+\infty}\frac{3}{7}\cdot x\delta(x)dx + \int_{-\infty}^{+\infty}\left(\begin{cases} 0, \text{если } x < 0, \\ \frac{x}{70}, \text{если } 0 < x < 40, \\ 0, \text{если } x > 40 \end{cases}\right)dx =$$

$$= \int_{-\infty}^{+\infty}\frac{3}{7}\cdot 0\cdot\delta(x)dx + \int_{-\infty}^{0} 0dx + \int_0^{40}\frac{x}{70}dx + \int_{40}^{+\infty} 0dx = 0 + 0 + \frac{1}{70}\cdot\left(\frac{x^2}{2}\right)\Bigg|_0^{40} + 0 = \frac{1}{70}\cdot\left(\frac{40^2}{2} - \frac{0^2}{2}\right) = \frac{20\cdot 40}{70} = \frac{80}{7}.$$

Этот результат можно получить, используя аналогию с центром тяжести. Найдём центр тяжести фигуры, содержащей массу 3/7, сосредоточенную в точке *x*=0, и массу 4/7, равномерно распределённую на промежутке [0;40). Заменим равномерно распределённую массу 4/7 такой же, но сосредоточенной в центре распределённой: точке *x*=20. А теперь получим математическое ожидание по обычной формуле для дискретной случайной величины: *EX*=0·(3/7)+20·(4/7)=80/7.

Распределение Пуассона

Исторически сложилось, что задача, которая сейчас будет рассмотрена, называется «задача о вызовах на телефонной станции». Условие такое: на телефонную станцию за некоторый промежуток времени в среднем поступает λ вызовов, причём их число в среднем прямо пропорционально длине промежутка времени. Вопрос: какова вероятность того, что за этот промежуток времени поступит ровно m вызовов?

Обозначим длину промежутка времени, введённого в условии числом T. Если ввести случайную величину X, равную числу вызовов, поступивших за время T, то из условия следует, что $EX=\lambda$. Именно это равенство соответствует информации о том, что за промежуток T в среднем поступает λ вызовов. Разобьём промежуток времени T на n частей так, что в каждой из таких частей не сможет поместиться более одного вызова. Случайная величина, равная числу вызовов за время T/n будет иметь распределение Бернулли. Обозначим за p её единственный параметр, равный вероятности того, что произойдет один (и единственный) вызов, за время T/n. Выше было выведено, что математическое ожидание случайной величины, имеющей распределение Бернулли равно $0\cdot(1-p)+1\cdot p=p$. Исходя из пропорциональности среднего числа вызовов длине промежутка, запишем равенство: $\frac{\lambda}{p}=\frac{T}{T/n}\Leftrightarrow\frac{\lambda}{p}=n\Leftrightarrow\lambda=np$. Вероятность того, что за промежуток времени T поступит ровно m вызовов, может быть интерпретирована, как предел при $n\to\infty$ вероятности того, что в серии из n независимых испытаний, с вероятностью успеха в одном испытании p, произойдёт ровно m успехов, причём произведение $np=\lambda$ оказывается постоянным числом, как это требуется в теореме, из которой следовала приближённая формула Пуассона. Устремив n к бесконечности, получим, что эта вероятность равна $\lim_{n\to\infty}P_n(m)=\frac{\lambda^m}{m!}\cdot e^{-\lambda}$, причём даже точно, а не приближённо. Это и есть ответ в задаче.

Распределение Пуассона с параметром λ имеет дискретная случайная величина, принимающая только целые неотрицательные значения с вероятностями: $P(X=m)=\frac{\lambda^m}{m!}\cdot e^{-\lambda}$. У неё бесконечное количество возможных значений. Таблица распределения такой дискретной случайной величины имеет вид:

x_i:	0	1	2	3	…	m	…
p_i:	$1/e^\lambda$	λ/e^λ	$\lambda^2/(2e^\lambda)$	$\lambda^3/(6e^\lambda)$	…	$\lambda^m/(m!e^\lambda)$	…

Убедимся в выполнении условия нормированности для распределения Пуассона: $\sum_{m=0}^{\infty}\frac{\lambda^m}{m!e^\lambda}=\frac{1}{e^\lambda}\sum_{m=0}^{\infty}\frac{\lambda^m}{m!}=\frac{1}{e^\lambda}\cdot e^\lambda=1$. Убедимся в том, что математическое ожидание для распределения Пуассона действительно равно λ: $EX=\sum_{m=0}^{\infty}\frac{m\lambda^m}{m!e^\lambda}=\sum_{m=1}^{\infty}\frac{m\lambda^m}{m!e^\lambda}=\frac{1}{e^\lambda}\sum_{m=1}^{\infty}\frac{\lambda^m}{(m-1)!}=\frac{\lambda}{e^\lambda}\sum_{m=1}^{\infty}\frac{\lambda^{m-1}}{(m-1)!}=\frac{\lambda}{e^\lambda}\sum_{n=0}^{\infty}\frac{\lambda^n}{n!}=\frac{\lambda}{e^\lambda}\cdot e^\lambda=\lambda$.

Если рассматривать временной промежуток, на котором в среднем происходит одно событие, то $EX=\lambda=1$. В этом случае вероятность того, что за этот период произойдёт ровно одно событие, равна вероятности того, что не произойдёт ни одного события, и эти вероятности равны $1/e$. На рисунке 25 показан график функции распределения для простейшего случая, когда $\lambda=1$.

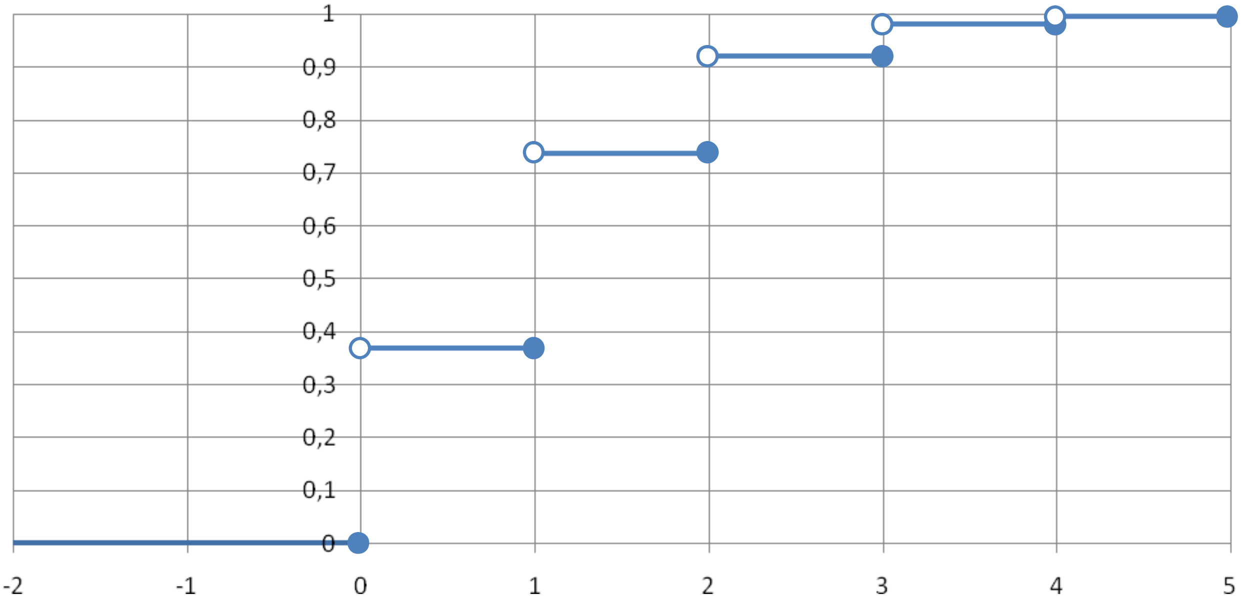

Рис. 25. График функции распределения Пуассоновской случайной величины (при λ=1).

Экспоненциальное распределение

Исторически сложилось, что задача, которая сейчас будет рассмотрена, называется «задача о времени ожидания автобуса». Условия такие: частота прихода автобусов (отношение среднего числа пришедших автобусов к интервалу времени, за которое они пришли) равна λ. Пассажир приходит на автобусную остановку в случайно выбранный момент времени. Каково математическое ожидание времени ожидания автобуса? Задачу нужно решить в двух постановках: а) автобусы ходят через равные промежутки времени; б) автобусы ходят хаотично и время прихода одного из них никак не зависит от времени прихода остальных.

Обозначим X случайную величину, равную времени ожидания автобуса.

Решение а). Найдём T, длину промежутка времени, через который ходят автобусы. Поскольку за время T приходит один автобус, то частота $\lambda=1/T \Leftrightarrow T=1/\lambda$. Это максимальное время ожидания автобуса при данной постановке задачи, а минимальное время ожидания это нуль. По условию пассажир (не знающий расписания автобусов) приходит на остановку в произвольный момент временного цикла, длиной $T=1/\lambda$. Поскольку вероятность прихода пассажира в любые промежутки времени равной длины, одинаковая, то время ожидания автобуса это случайная величина, равномерно распределённая на промежутке от 0 до $1/\lambda$. Её математическое ожидание равно $EX=\dfrac{0+1/\lambda}{2}=\dfrac{1}{2\lambda}$.

Решение б). Вероятность $P(X<x)$ уехать до времени $x\geq 0$ это вероятность события, противоположного событию «за время x не будет ни одного автобуса», то есть событию «за время x придёт ровно нуль автобусов». А такую вероятность можно вычислить, используя распределение Пуассона. По условию среднее количество автобусов, пришедших за промежуток времени длиной x равно λx. Учитывая правило нахождения вероятности противоположного события, получим для $x\geq 0$ функцию распределения этой случайной величины $F(x)=P(X<x)=1-P(\textit{придёт нуль автобусов})=1-\dfrac{(\lambda x)^0}{0!}e^{-\lambda x}=1-\dfrac{1}{1}e^{-\lambda x}=1-e^{-\lambda x}$. Если x отрицательно, то $F(x)=P(X<x)=0$, поскольку ждать отрицательное время невозможно. Таким образом, функция распределения равна кусочно-заданной функции:

$F(x)=\begin{cases}0, \text{ если } x<0,\\ 1-e^{-\lambda x}, \text{ если } x\ge 0.\end{cases}$ Производная от неё будет плотностью вероятности: $f(x)=\begin{cases}0, \text{ если } x<0,\\ \lambda e^{-\lambda x}, \text{ если } x>0.\end{cases}$ Математическое ожидание находится по общей формуле: $EX=\int\limits_{-\infty}^{+\infty} xf(x)dx=\int\limits_{-\infty}^{0} xf(x)dx+\int\limits_{0}^{+\infty} xf(x)dx=\int\limits_{-\infty}^{0} x0dx+\int\limits_{0}^{+\infty} x\lambda e^{-\lambda x}dx=0+\lim\limits_{b\to+\infty}\int\limits_{0}^{b} x\lambda e^{-\lambda x}dx$. Этот интеграл возьмём по частям: $u=\lambda x\Rightarrow du=\lambda\cdot dx$; $dv=e^{-\lambda x}dx\Rightarrow v=-\frac{1}{\lambda}e^{-\lambda x}$. Тогда

$$EX=\lim_{b\to+\infty}\left(\left.\left(-\lambda x\cdot\frac{1}{\lambda}e^{-\lambda x}\right)\right|_0^b-\int\limits_0^b-\lambda\cdot\frac{1}{\lambda}e^{-\lambda x}dx\right)=\lim_{b\to+\infty}\left(\left.\left(-xe^{-\lambda x}\right)\right|_0^b+\int\limits_0^b e^{-\lambda x}dx\right)$$

$$=\lim_{b\to+\infty}\left(-be^{-\lambda b}+0-\left.\left(\frac{1}{\lambda}e^{-\lambda x}\right)\right|_0^b\right)=-\lim_{b\to+\infty}\frac{b}{e^{\lambda b}}-\lim_{b\to+\infty}\left(\frac{1}{\lambda}e^{-\lambda b}-\frac{1}{\lambda}e^0\right)=-\lim_{b\to+\infty}\frac{1}{\lambda e^{\lambda b}}-\left(0-\frac{1}{\lambda}\right)=$$

$=-0+\frac{1}{\lambda}=\frac{1}{\lambda}$. Ответ оказался вдвое больше, чем в задаче, где автобусы ходили строго через равные промежутки времени по расписанию. Это для того же количества автобусов, но при полной анархии. Заметим ещё, что при хаотическом движении автобусов, вероятность сесть в переполненный автобус больше, чем в свободный, поскольку промежуток времени перед его приходом больше и в нём оказывается больше пассажиров.

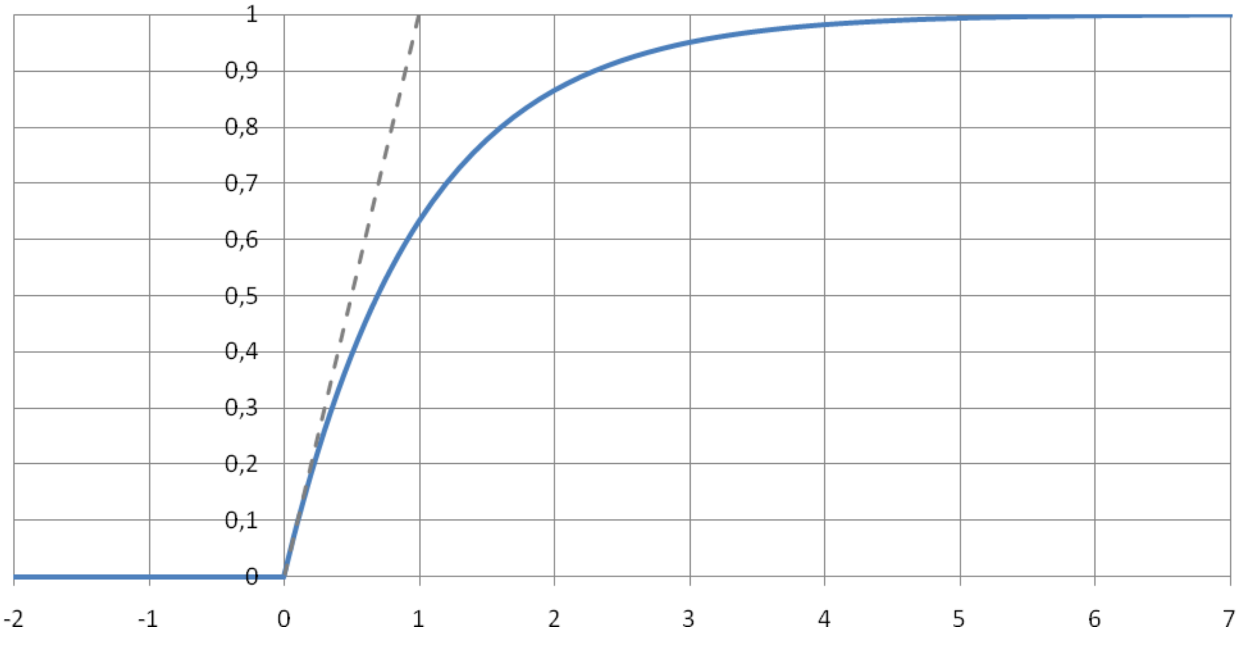

Рис. 26. График функции экспоненциального распределения при λ=1.

Случайная величина с функцией распределения $F(x)=\begin{cases}0, \text{ если } x<0,\\ 1-e^{-\lambda x}, \text{ если } x\ge 0.\end{cases}$ называется экспоненциальной (или показательной) с параметром λ. Покажем для неё выполнение условия нормированности:

$$\int_{-\infty}^{+\infty} f(x)dx = \int_{-\infty}^{0} f(x)dx + \int_{0}^{+\infty} f(x)dx = \int_{-\infty}^{0} x0dx + \int_{0}^{+\infty} \lambda e^{-\lambda x}dx = 0 + \lim_{b\to+\infty}\int_{0}^{b} \lambda e^{-\lambda x}dx =$$

$= \lim\limits_{b\to+\infty}\left(-\lambda\cdot\frac{1}{\lambda}\cdot e^{-\lambda x}\right)\Big|_0^b = \lim\limits_{b\to+\infty}\left(-e^{-\lambda b}+1\right) = 1$. График её функции распределения при λ=1 показан на рис. 26. Пунктиром нарисован фрагмент графика функции распределения равномерно распределённой случайной величины, равной времени ожидания автобуса для случая, когда автобусы приходят с той же интенсивностью, но строго через равные промежутки времени. Интересно, что пунктирная линия оказывается касательной к графику в точке x=0. То есть, при малых положительных x эти графики неразличимы. Практически это означает, что если пассажиру повезло, и он мало времени ждал автобуса, то он не сможет догадаться, как ходят автобусы, по расписанию или хаотично. На рис. 27 показан график плотности вероятности экспоненциальной случайной величины при λ=1.

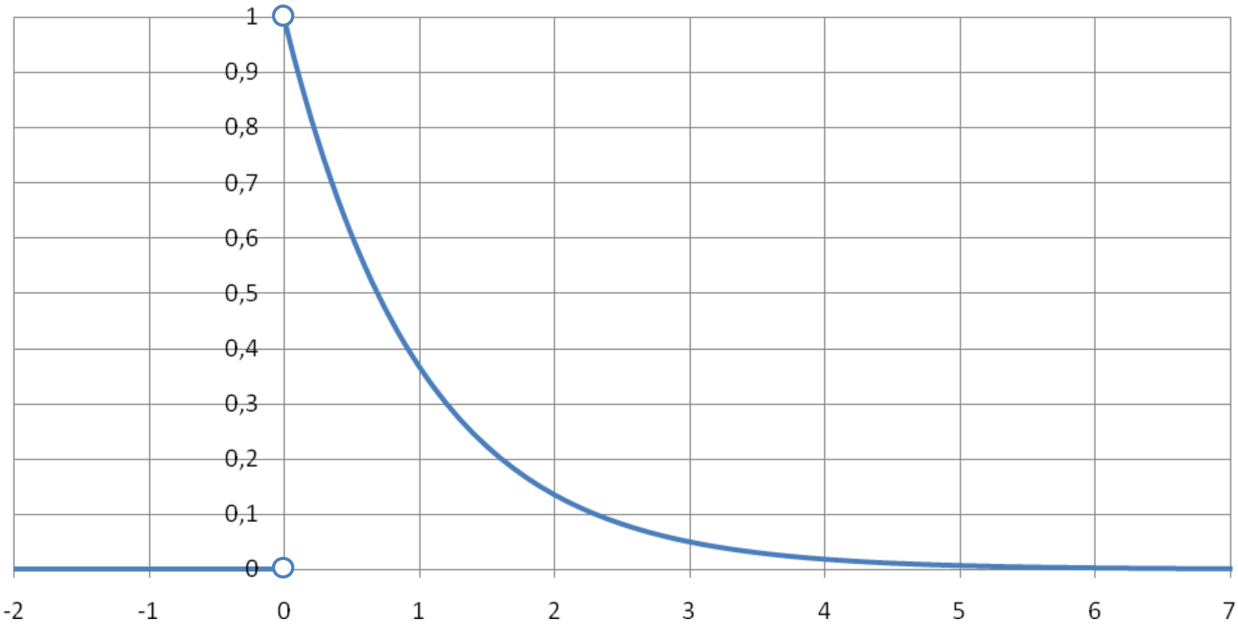

Рис. 27. График плотность вероятности экспоненциального распределения при λ=1.

Распределение Коши

Организуем любым доступным способом равномерно распределённую на промежутке $(-\pi/2;\pi/2)$ случайную величину T. Её функция распределения будет иметь вид

$$G(x) = \begin{cases} 0, \text{если } x \le -\dfrac{\pi}{2}, \\ \dfrac{1}{2} + \dfrac{x}{\pi}, \text{если } -\dfrac{\pi}{2} < x \le \dfrac{\pi}{2}, \\ 1, \text{если } x > \dfrac{\pi}{2}. \end{cases}$$

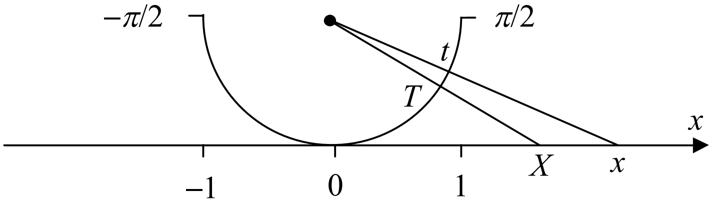

Рис. 28. Определение распределения Коши.

Числа промежутка $(-\pi/2;\pi/2)$ нанесём на полуокружность единичного радиуса (см. рис. 28). Построим числовую прямую, касающуюся полуокружности в точке нуль, и имеющую точку касания своим началом. После того, как в эксперименте получится значение случайной величины T, проведём луч из центра через полученную точку полуокружности до пересечения с касательной. При этом образуется число X, являющееся значением новой случайной величины, распределение которой называется распределением Коши. Найдём её функцию распределения $F(x)$.

$F(x)=P(X<x)=P(T<t)=P(T<\operatorname{arctg}x)=G(\operatorname{arctg}x)=\dfrac{1}{2}+\dfrac{\operatorname{arctg}x}{\pi}$. Последний переход объясняется тем, что $-\dfrac{\pi}{2}<\operatorname{arctg}x<\dfrac{\pi}{2}$. График функции распределения показан на рисунке 29.

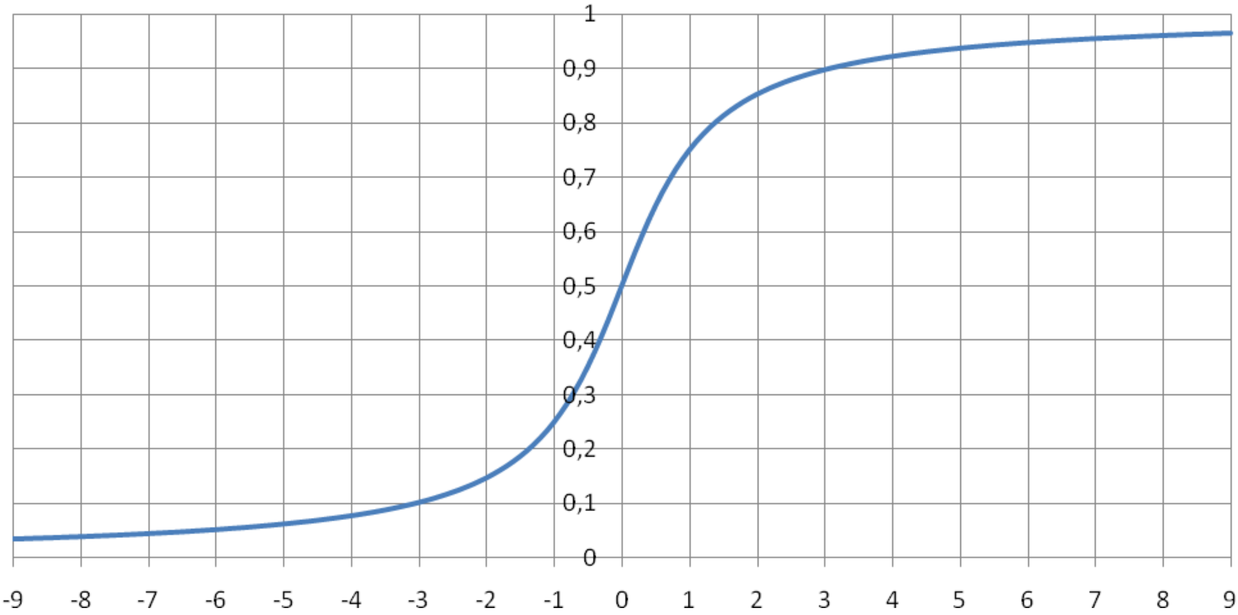

Рис. 29. Функция распределения случайной величины Коши.

Плотность вероятности распределения Коши равна производной от найденной функции распределения $f(x)=F'(x)=\left(\dfrac{1}{2}+\dfrac{\operatorname{arctg}x}{\pi}\right)'=\dfrac{1}{\pi\left(1+x^2\right)}$. Убедимся в выполнении условия нормированности: $\int\limits_{-\infty}^{+\infty}f(x)dx=\int\limits_{-\infty}^{+\infty}\dfrac{dx}{\pi\left(1+x^2\right)}=\dfrac{2}{\pi}\int\limits_{0}^{+\infty}\dfrac{dx}{1+x^2}=\dfrac{2}{\pi}\lim\limits_{b\to+\infty}\int\limits_{0}^{b}\dfrac{dx}{1+x^2}=\dfrac{2}{\pi}\lim\limits_{b\to+\infty}\operatorname{arctg}x\Big|_0^b=$

$=\dfrac{2}{\pi}\lim\limits_{b\to+\infty}\left(\operatorname{arctg}b-0\right)=\dfrac{2}{\pi}\cdot\dfrac{\pi}{2}=1$. График плотности вероятности показан на рисунке 30.

Докажем, что у распределения Коши нет математического ожидания, для чего убедимся в расходимости несобственного интеграла:

$\int\limits_{-\infty}^{+\infty}\left|xf(x)\right|dx=\int\limits_{-\infty}^{+\infty}\dfrac{|x|\,dx}{\pi\left(1+x^2\right)}=\dfrac{2}{\pi}\int\limits_{0}^{+\infty}\dfrac{x\,dx}{1+x^2}$. Сделаем замену переменных: $t=1+x^2\Rightarrow dt=2x\cdot dx$, тогда если $x=0$, то $t=1$, а если $x\to+\infty$, то $t\to+\infty$, поэтому

$\int\limits_{-\infty}^{+\infty}\left|xf(x)\right|dx=\dfrac{1}{\pi}\int\limits_{1}^{+\infty}\dfrac{dt}{t}=\dfrac{1}{\pi}\lim\limits_{b\to+\infty}\int\limits_{1}^{b}\dfrac{dt}{t}=\dfrac{1}{\pi}\lim\limits_{b\to+\infty}\ln t\Big|_1^b=\dfrac{1}{\pi}\lim\limits_{b\to+\infty}\left(\ln b-\ln 1\right)=+\infty$. Несобственный интеграл расходится, поэтому математического ожидания распределения Коши не существует, хотя графики функции распределения и плотности вероятности качественно напоминают соответствующие графики для нормального распределения. У графика плотности вероятности так же есть перегибы $\left(\text{при } x=\pm\dfrac{1}{\sqrt{3}}\right)$. Только эти функции значительно медленнее (по сравнению с соответствующими им графиками нормального распределения) стремятся к своим пределам на бесконечности.

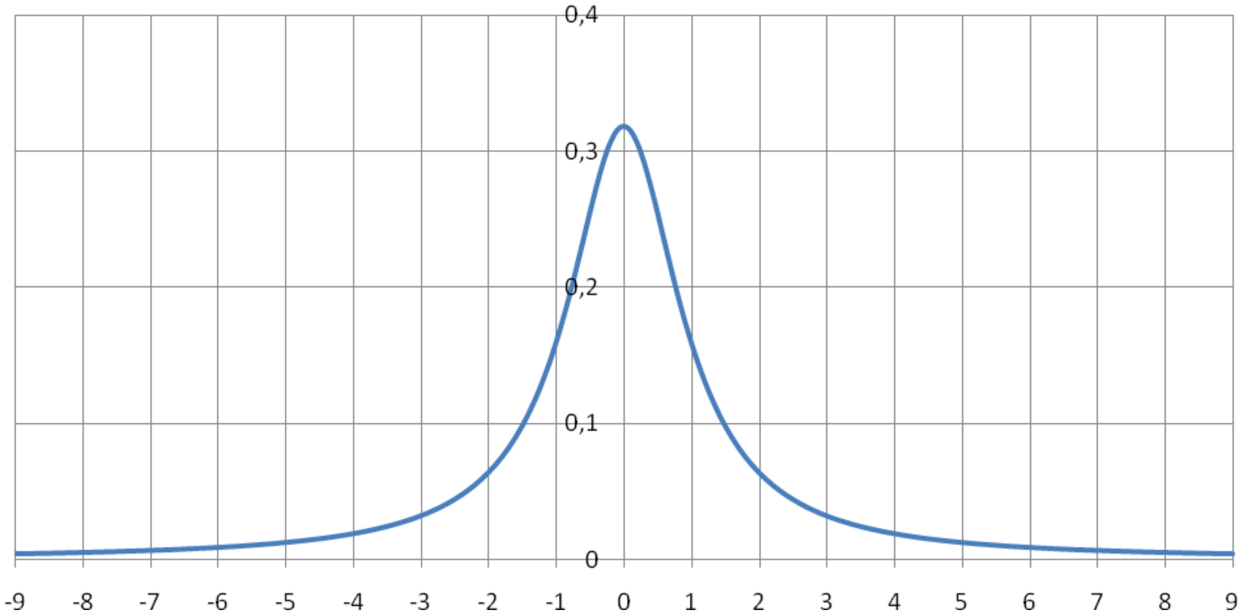

Рис. 30. Плотность вероятности распределения Коши.

Преобразование случайной величины

Пусть борелевская (ответ на вопрос: «какие именно функции называются борелевскими?», не входит в курс теории вероятностей) функция $R(x)$ определена на множестве значений случайной величины X. Определим случайную величину $Y=R(X)$. Другими словами, в эксперименте получается значение случайной величины X, а потом с помощью функци $R(x)$ пересчитывается *наведённая* случайная величина Y. Такие случайные величины автоматически принимают значения одновременно, и оказываются определёнными на одном вероятностном пространстве. Если X является дискретной случайной величиной, то случайная величина Y тоже будет дискретной, причём с возможными значениями $y_i=R(x_i)$, которые реализуются с теми же вероятностями: $P(Y=y_i)=P(X=x_i)=p_i$ (разумеется, если некоторые возможные значения y_i будут совпадать друг с другом, то повторяющиеся значения y_i нужно будет изьять, а соответствующие вероятности p_i сложить). Таблица распепеределения случайных величин X и Y имеет вид:

x_i:	x_1	x_2	x_3	…	x_n	…
y_i:	$R(x_1)$	$R(x_2)$	$R(x_3)$	…	$R(x_n)$	…
p_i:	p_1	p_2	p_3	…	p_n	…

Математическое ожидание случайной величины Y находится по обычной формуле:

$EY=\sum_i y_i p_i=\sum_i R(x_i)p_i$. Заметим, что эта формула отличается от формулы для EX тем, что x_i формально заменено на $R(x_i)$. Примем без доказательства, что аналогичной заменой x на $R(x)$, где $R(x)$ борелевская функция, можно получить формулу для нахождения математического ожидания не только дискретной величины и $EY=\int_{-\infty}^{+\infty} R(x)f(x)dx$.

Независимость двух случайных величин

Пусть две случайные величины X и Y определены на одном вероятностном пространстве. Это означает, что они принимают возможные значения одновременно. Случайные величины X и Y называются независимыми, если для любых чисел x и y независимы события $X<x$ и $Y<y$, то есть если выполняется равенство $P((X<x)\cap(Y<y))=P(X<x)\cdot P(Y<y)$. Если функциями распределения случайных величин X и Y являются функции $F(x)$ и $G(x)$ соответственно, то это равенство может быть продолжено: $P((X<x)\cap(Y<y))=F(x)\cdot G(y)$.

Докажем свойства независимых случайных величин.

1. Вырожденная случайная величина независима с любой другой случайной величиной. Пусть X — имеет вырожденное распределение и всегда принимает значение c, то есть $P(X=c)=1$. Тогда при $x\leq c$ событие $X<x$ невозможное, а при $x>c$ достоверное. В любом случае, оно независимо с произвольным событием. Поэтому вырожденная случайная величина независима с любой случайной величиной.
2. Практически очевидно, что если X и Y независимые случайные величины, то будут независимы и случайные величины X и $Z=R(Y)$, поскольку при образовании значения случайной величины Z значение X никак во внимание не принимается. Свойство примем без строгого доказательства.
3. Правило произведения для промежутков: если X и Y независимые случайные величины, а функции $F(x)$ и $G(x)$ являются их функциями распределения, соответственно, то $P((a\leq X<b)\cap(c\leq Y<d))=P(a\leq X<b)\cdot P(c\leq Y<d)=(F(b)-F(a))\cdot(G(d)-G(c))$. Для доказательства разобьём событие $(X<b)\cap(c\leq Y<d)$ на два несовместных события: $(X<b)\cap(c\leq Y<d)=((X<a)\cup(a\leq X<b))\cap(c\leq Y<d)=((X<a)\cap(c\leq Y<d))\cup((a\leq X<b)\cap(c\leq Y<d))$. Тогда по аксиоме сложения $P((X<b)\cap(c\leq Y<d))=$ $=P((X<a)\cap(c\leq Y<d))+P((a\leq X<b)\cap(c\leq Y<d))\Leftrightarrow P((a\leq X<b)\cap(c\leq Y<d))=$ $=P((X<b)\cap(c\leq Y<d))-P((X<a)\cap(c\leq Y<d))$. Теперь разобьём событие $(X<b)\cap(Y<d)$ на два несовместных события: $(X<b)\cap(Y<d)=(X<b)\cap((Y<c)\cup(c\leq Y<d))=$ $=((X<b)\cap(Y<c))\cup((X<b)\cap(c\leq Y<d))$. Тогда по аксиоме сложения $P((X<b)\cap(Y<d))=$ $=P((X<b)\cap(Y<c))+P((X<b)\cap(c\leq Y<d))\Leftrightarrow P((X<b)\cap(c\leq Y<d))=$ $=P((X<b)\cap(Y<d))-P((X<b)\cap(Y<c))=F(b)\cdot G(d)-F(b)\cdot G(c)=F(b)\cdot(G(d)-G(c))$. Аналогично (заменяя в полученном равенстве литеру b на литеру a) можно получить: $P((X<a)\cap(c\leq Y<d))=P((X<a)\cap(Y<d))-P((X<a)\cap(Y<c))=F(a)\cdot(G(d)-G(c))$. Использовав эти два равенства получим $P((a\leq X<b)\cap(c\leq Y<d))=$ $=P((X<b)\cap(Y<d))-P((X<b)\cap(Y<c))-(P((X<a)\cap(Y<d))-P((X<a)\cap(Y<c)))=$ $=P((X<b)\cap(Y<d))-P((X<b)\cap(Y<c))-P((X<a)\cap(Y<d))+P((X<a)\cap(Y<c))$. Полученное соотношение верно и для зависимых случайных величин X и Y, а при условии независимости равенство можно продолжить: $P((a\leq X<b)\cap(c\leq Y<d))=$ $=F(b)\cdot G(d)-F(b)\cdot G(c)-F(a)\cdot G(d)+F(a)\cdot G(c)=$ $=F(b)\cdot(G(d)-G(c))-F(a)\cdot(G(d)-G(c))=(F(b)-F(a))\cdot(G(d)-G(c))=P(a\leq X<b)\cdot P(c\leq Y<d)$. Правило доказано.
4. Правило произведения для возможных значений: если X и Y независимые случайные величины, то $P((X=x_0)\cap(Y=y_0))=P(X=x_0)\cdot P(Y=y_0)$. Заметим, что если $P(X=x_0)=0$, что справедливо, например, если X непрерывная случайная величина или если x_0 не является возможным значением дискретной случайной величины, то в этом случае по монотонности вероятности из того, что $((X=x_0)\cap(Y=y_0))\subset(X=x_0)$ следует $P((X=x_0)\cap(Y=y_0))\leq P(X=x_0)=0\Rightarrow P((X=x_0)\cap(Y=y_0))=0$. После чего $P((X=x_0)\cap(Y=y_0))=P(X=x_0)\cdot P(Y=y_0)\Leftrightarrow 0=0\cdot P(Y=y_0)\Leftrightarrow 0=0$ оказывается верным, но бесполезным равенством. Таким образом, правило произведения для возможных значений оказывается содержательным только в том случае, когда x_0 и y_0 являются возможными отдельно стоящими значениями дискретных (или смешанных) случайных величин X и Y. Докажем это правило только для дискретных случайных

величин. Выберем числа a, b, c и d так, чтобы промежутки $(a;b)$ и $(c;d)$ содержали только по одному возможному значению: $x_0 \in (a;b)$, $y_0 \in (c;d)$. Тогда событие $(X=x_0)=(a \le X<b)$, а событие $(Y=y_0)=(c \le Y<d)$. Поэтому $P((X=x_0)\cap(Y=y_0))=$ $=P((a\le X<b)\cap(c\le Y<d))=P(a\le X<b)\cdot P(c\le Y<d)=P(X=x_0)\cdot P(Y=y_0)$. Для смешанных случайных величин правило примем без доказательства.

Независимость трёх и более случайных величин

Если случайных величин три или больше, то, как и для событий, различают два вида независимости: попарную и взаимную.

Случайные величины X_1, X_2, X_3, … , X_n (определённые на одном вероятностном пространстве) называются попарно независимыми, если для любых n чисел x_1, x_2, x_3, … , x_n события $X_1<x_1$, $X_2<x_2$, $X_3<x_3$, … , $X_n<x_n$ будут попарно независимыми, то есть для любых $i\neq j$ верно, что $P((X_i<x_i)\cap(X_j<x_j))=P(X_i<x_i)\cdot P(X_j<x_j)$. В частности, три случайных величины X, Y и Z будут попарно независимыми, если для любых трёх чисел x, y, z верно, что

$$\begin{cases} P((X<x)\cap(Y<y))=P(X<x)\cdot P(Y<y), \\ P((X<x)\cap(Z<z))=P(X<x)\cdot P(Z<z), \\ P((Y<y)\cap(Z<z))=P(Y<y)\cdot P(Z<z). \end{cases}$$

Случайные величины X_1, X_2, X_3, … , X_n (определённые на одном вероятностном пространстве) называются взаимно независимыми (или независимыми в совокупности), если для любого набора из этих n случайных величин события $X_1<x_1$, $X_2<x_2$, $X_3<x_3$, … , $X_n<x_n$ будут взаимно независимыми. В частности, три случайных величины X, Y и Z будут попарно независимыми, если для любых трёх чисел x, y, z верно, что

$$\begin{cases} P((X<x)\cap(Y<y))=P(X<x)\cdot P(Y<y), \\ P((X<x)\cap(Z<z))=P(X<x)\cdot P(Z<z), \\ P((Y<y)\cap(Z<z))=P(Y<y)\cdot P(Z<z), \\ P((X<x)\cap(Y<y)\cap(Z<z))=P(X<x)\cdot P(Y<y)\cdot P(Z<z). \end{cases}$$

Из этих двух определений следует, что если случайные величины взаимно независимые, то они обязательно попарно независимые, поскольку аналогичное утверждение верно для событий. Полная аналогия также в том, что употребление термина *независимые случайные величины* без уточнения как именно независимые, означает взаимную независимость.

Для любого набора из взаимно независимых случайных величин справедливы правила произведения и для промежутков и для возможных значений.

Свойства математического ожидания

1. Математическое ожидание константы равно этой константе: $EC=C\cdot 1=C$.
2. Сомножитель-константу можно выносить за знак математического ожидания. Пусть случайная величина $Y=aX$, тогда функция $R(x)$ из предыдущего пункта равна $R(x)=a\cdot x$, поэтому: $E(aX)=EY=\int_{-\infty}^{+\infty} axf(x)dx = a\int_{-\infty}^{+\infty} xf(x)dx = a\cdot EX$.
3. Математическое ожидание суммы случайных величин, определённых на одном вероятностном пространстве, равно сумме их математических ожиданий. Докажем это утверждение для дискретных случайных величин X и Y. Если $P(X=x_i)=u_i$, то таблица распределения случайной величины X имеет вид:

x_i:	x_1	x_2	x_3	…	x_i	…
$P(X=x_i)$:	u_1	u_2	u_3	…	u_i	…

Если $P(Y=y_j)=v_j$, то таблица распределения для случайной величины Y:

y_i:	y_1	y_2	y_3	…	y_j	…
$P(Y=y_i)$:	v_1	v_2	v_3	…	v_j	…

Поскольку случайные величины X и Y одновременно принимают свои возможные значения, то существует случайная величина $Z=X+Y$, возможными значениями которой являются числа $z_{ij}=x_i+y_j$, которые она принимает с вероятностями $P(Z=z_{ij})=P((X=x_i)\cap(Y=y_j))=p_{ij}$. Построим таблицу, в которую внесены возможные значения $z_{ij}=x_i+y_j$ случайной величины Z:

z_{ij}:	y_1	y_2	y_3	…	y_j	…
x_1	x_1+y_1	x_1+y_2	x_1+y_3	…	x_1+y_j	…
x_2	x_2+y_1	x_2+y_2	x_2+y_3	…	x_2+y_j	…
x_3	x_3+y_1	x_3+y_2	x_3+y_3	…	x_3+y_j	…
…	…	…	…	…	…	…
x_i	x_i+y_1	x_i+y_2	x_i+y_3	…	x_i+y_j	…
…	…	…	…	…	…	…

Актуально также составить таблицу, в которой записаны вероятности одновременной реализации событий $P(Z=z_{ij})=P((X=x_i)\cap(Y=y_j))=p_{ij}$:

$P(Z=z_{ij})$:		v_1	v_2	v_3	…	v_j	…
		y_1	y_2	y_3	…	y_j	…
u_1	x_1	p_{11}	p_{12}	p_{13}	…	p_{1j}	…
u_2	x_2	p_{21}	p_{22}	p_{23}	…	p_{2j}	…
u_3	x_3	p_{31}	p_{32}	p_{33}	…	p_{3j}	…
…	…	…	…	…	…	…	…
u_i	x_i	p_{i1}	p_{i2}	p_{i3}	…	p_{ij}	…
…	…	…	…	…	…	…	…

Вероятности p_{ij} должны быть неотрицательными и удовлетворять условию нормированности: $\sum_{i,j} p_{ij}=1$. Заметим, что $u_i=P(X=x_i)=P((X=x_i)\cap\Omega)=$

$=P((X=x_i)\cap((Y=y_1)\cup(Y=y_2)\cup(Y=y_3)\cup\ldots))=$

$=P(((X=x_i)\cap(Y=y_1))\cup((X=x_i)\cap(Y=y_2))\cup((X=x_i)\cap(Y=y_3))\cup\ldots)=$

$=P((X=x_i)\cap(Y=y_1))+P((X=x_i)\cap(Y=y_2))+P((X=x_i)\cap(Y=y_3))+\ldots=p_{i1}+p_{i2}+p_{i3}+\ldots=\sum_j p_{ij}$.

Аналогично: $v_j=P(Y=y_j)=P(\Omega\cap(Y=y_j))=P(((X=x_1)\cup(X=x_2)\cup(X=x_3)\cup\ldots)\cap(Y=y_j))=$
$=P(((X=x_1)\cap(Y=y_j))\cup((X=x_2)\cap(Y=y_j))\cup((X=x_3)\cap(Y=y_j))\cup\ldots)=P((X=x_1)\cap(Y=y_j))+$
$+P((X=x_2)\cap(Y=y_j))+P((X=x_3)\cap(Y=y_j))+\ldots=p_{1j}+p_{2j}+p_{3j}+\ldots=\sum_i p_{ij}$. Найдём

$$E(X+Y)=EZ=\sum_{i,j} z_{ij}p_{ij}=\sum_{i,j}(x_i+y_j)p_{ij}=\sum_{i,j}(x_ip_{ij}+y_jp_{ij})=\sum_{i,j}x_ip_{ij}+\sum_{i,j}y_jp_{ij}=$$

$$=\sum_i\sum_j x_ip_{ij}+\sum_j\sum_i y_jp_{ij}=\sum_i x_i\sum_j p_{ij}+\sum_j y_j\sum_i p_{ij}=\sum_i x_iu_i+\sum_j y_jv_j=EX+EY.$$

4. Предыдущие два свойства можно объединить в свойство линейности: математическое ожидание линейной комбинации случайных величин (определённых на одном вероятностном пространстве) равно линейной комбинации (с теми же коэффициентами) математических ожиданий этих случайных величин: $E(aX+bY)=a\cdot EX+b\cdot EY$.

5. В частности, $E(X+C)=EX+C$.
6. Частным случаем линейности также является то, что математическое ожидание разности случайных величин (определённых на одном вероятностном пространстве) равно разности их математических ожиданий: $E(X-Y)=EX-EY$.
7. Математическое ожидание произведения <u>независимых</u> случайных величин (определённых на одном вероятностном пространстве) равно произведению их математических ожиданий. Это утверждение, также как и третье свойство, докажем только для дискретных случайных величин X и Y, таблицы распределения которых имеют вид:

x_i:	x_1	x_2	x_3	…	x_i	…
$P(X=x_i)$:	u_1	u_2	u_3	…	u_i	…

и

y_i:	y_1	y_2	y_3	…	y_j	…
$P(Y=y_i)$:	v_1	v_2	v_3	…	v_j	…

Пусть случайная величина $Z=X\cdot Y$. Тогда её возможные значения равны $z_{ij}=x_i\cdot y_j$, и случайная величина Z принимает их с вероятностями $P(Z=z_{ij})=p_{ij}$. Из независимости случайных величин X и Y следуют равенства: $P(Z=z_{ij})=P((X=x_i)\cap(Y=y_j))=P(X=x_i)\cdot P(Y=y_j)=u_i\cdot v_j$. Построим таблицу, в которую внесены возможные значения $z_{ij}=x_i\cdot y_j$ случайной величины Z:

z_{ij}:	y_1	y_2	y_3	…	y_j	…
x_1	$x_1\cdot y_1$	$x_1\cdot y_2$	$x_1\cdot y_3$	…	$x_1\cdot y_j$	…
x_2	$x_2\cdot y_1$	$x_2\cdot y_2$	$x_2\cdot y_3$	…	$x_2\cdot y_j$	…
x_3	$x_3\cdot y_1$	$x_3\cdot y_2$	$x_3\cdot y_3$	…	$x_3\cdot y_j$	…
…	…	…	…	…	…	…
x_i	$x_i\cdot y_1$	$x_i\cdot y_2$	$x_i\cdot y_3$	…	$x_i\cdot y_j$	…
…	…	…	…	…	…	…

Составим таблицу, в которой записаны вероятности одновременной реализации событий $P(Z=z_{ij})=P((X=x_i)\cap(Y=y_j))=u_i\cdot v_j$:

$P(Z=z_{ij})$:		v_1	v_2	v_3	…	v_j	…
		y_1	y_2	y_3	…	y_j	…
u_1	x_1	$u_1\cdot v_1$	$u_1\cdot v_2$	$u_1\cdot v_3$	…	$u_1\cdot v_j$	…
u_2	x_2	$u_2\cdot v_1$	$u_2\cdot v_2$	$u_2\cdot v_3$	…	$u_2\cdot v_j$	…
u_3	x_3	$u_3\cdot v_1$	$u_3\cdot v_2$	$u_3\cdot v_3$	…	$u_3\cdot v_j$	…
…	…	…	…	…	…	…	…
u_i	x_i	$u_i\cdot v_1$	$u_i\cdot v_2$	$u_i\cdot v_3$	…	$u_i\cdot v_j$	…
…	…	…	…	…	…	…	…

Поэтому: $$E(XY)=EZ=\sum_{i,j} z_{ij}p_{ij}=\sum_{i,j}(x_iy_j)(u_iv_j)=\sum_j\sum_i x_iy_ju_iv_j=\sum_j\left(y_jv_j\sum_i x_iu_i\right)=$$
$$=\left(\sum_i x_iu_i\right)\cdot\left(\sum_j y_jv_j\right)=EX\cdot EY.$$

8. Математическое ожидание неотрицательной случайной величины тоже неотрицательное. Чтобы доказать это заметим, что, если случайная величина X не может принимать отрицательные значения, то $P(X<0)=0\Leftrightarrow F(0)=0$. А поскольку $\lim\limits_{x\to-\infty}F(x)=0$ и функция распределения $F(x)$ неубывающая, то $F(x)=0$ при любом

отрицательном x. Плотность вероятности равна производной функции распределения, а производная константы (в том числе и нуля) равна нулю. Поэтому при $x<0$ плотность вероятности $f(x)=0$. И математическое ожидание $EX = \int_{-\infty}^{+\infty} xf(x)dx = \int_{-\infty}^{0} xf(x)dx + \int_{0}^{+\infty} xf(x)dx = \int_{-\infty}^{0} x \cdot 0 \cdot dx + \int_{0}^{+\infty} xf(x)dx = \int_{0}^{+\infty} xf(x)dx \geq 0$, как интеграл от неотрицательной функции.

Математическое ожидание биномиального распределения

Пусть X_i случайная величина, имеющая распределение Бернулли с параметром p это число успехов в i-ом испытании. Тогда случайную величину Y, имеющую биномиальное распределение с параметрами n и p можно трактовать как $Y = \sum_{i=1}^{n} X_i$. Следовательно, $EY = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} EX_i = \sum_{i=1}^{n} p = np$.

Дисперсия и среднеквадратическое отклонение

Дисперсией DX случайной величины X, имеющей математическое ожидание EX называется математическое ожидание случайной величины $Y=(X-EX)^2$. Если считать, что $Y=R(X)$, то $R(x)=(x-EX)^2$. Согласно определению $DX = E\left((X-EX)^2\right) = \int_{-\infty}^{+\infty} (x-EX)^2 f(x)dx$.

Для дискретной случайной величины справедливо $DX = \sum_i (x_i - EX)^2 p_i$. Почему формула для дисперсии именно такая? Синонимом слова дисперсия является слово «рассеяние». Формула для дисперсии напоминает формулу для нахождения расстояния между двумя точками. Это расстояние между многомерной точкой с координатами x_i и точкой, у которой все координаты равны друг другу и равны EX (p_i это масштабные весовые коэффициенты координат по разным осям). Поэтому дисперсия характеризует разбросанность случайной величины вокруг её математического ожидания. Чем меньше эта разбросанность, тем точнее математическое ожидание характеризует случайную величину.

Формула для дисперсии отличается от формулы для расстояния между двумя точками и отсутствием квадратного корня. Для квадратного корня из дисперсии существует название: *среднеквадратическое* (или среднее квадратическое) *отклонение*. Оно обозначается $\sigma(X) = \sqrt{DX}$. Среднеквадратическое отклонение удобно тем, что оно имеет размерность случайной величины. Среднеквадратическое отклонение, так же как и дисперсия, характеризует разбросанность случайной величины вокруг её математического ожидания.

Найдем дисперсию случайной величины, порождённой бросанием игральной кости. Составим таблицу распределения вспомогательной случайной величины Y (напомним, что $EX=3{,}5$):

x_i:	1	2	3	4	5	6
x_i-EX	–2,5	–1,5	–0,5	0,5	1,5	2,5
$y_i=(x_i-EX)^2$	6,25	2,25	0,25	0,25	2,25	6,25
p_i:	1/6	1/6	1/6	1/6	1/6	1/6

DX=(6,25·(1/6)+2,25·(1/6)+0,25·(1/6))·2=$2\frac{11}{12}$, $\sigma(X)\approx1{,}707825$. Что эти числа обозначают сейчас не понятно, поскольку сравнивать эту разбросанность пока не с чем.

Найдём дисперсию случайной величины, имеющей распределение Бернулли с параметром p. Вспомним, что $EX=p$. Составим таблицу распределения величины $Y=(X-EX)^2$:

x_i:	0	1
x_i-EX	$-p$	$1-p$
$y_i=(x_i-EX)^2$	p^2	$(1-p)^2$
p_i:	$1-p$	p

Тогда $DX=p^2(1-p)+(1-p)^2p=p(1-p)(p-(1-p))=$ $=p(1-p)$. А среднеквадратическое отклонение, соответственно, равно $\sigma(X)=\sqrt{p(1-p)}$. Построим график этой функции, учитывая, что $p\in[0;1]$. Для этого возведём уравнение в квадрат, запомнив, что $\sigma\geq0$: $\sigma^2=p-p^2\Leftrightarrow$ $\Leftrightarrow\sigma^2+p^2-p=0\Leftrightarrow\sigma^2+p^2-p+(¼)=¼\Leftrightarrow$ $\Leftrightarrow\sigma^2+(p-½)^2=(½)^2$. Получится уравнение окружности с центром в точке (½;0) и радиусом ½. С учётом того, что $\sigma\geq0$, графиком будет являться верхняя половина этой окружности (см. рис. 31). Из графика следует, что наибольшую разбросанность случайная величина Бернулли достигает при p=½ (в симметричном случае), а наименьшую (нуля) при полной определённости, то есть когда является не случайной и всегда равна одному и тому же числу (нулю или единице, при p=0 и p=1, соответственно).

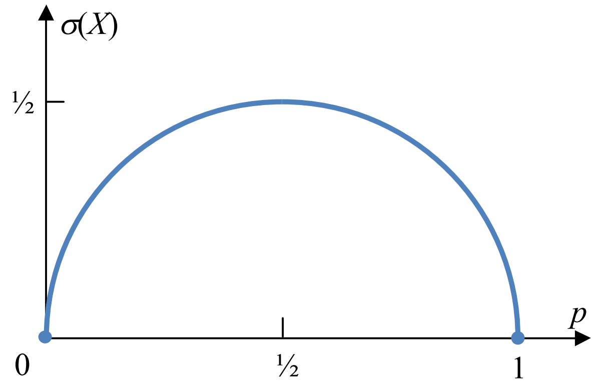

Рис. 31. Среднеквадратическое отклонение случайной величины Бернулли.

Найдём дисперсию дискретной случайной величины, распределённой по закону Пуассона с параметром λ. Напомним, что это величина, принимающая значения 0,1,2,3,… с вероятностями $P(X=m)=\frac{\lambda^m}{m!}e^{-\lambda}$, и математическим ожиданием $EX=\lambda$.

$$DX=\sum_{m=0}^{\infty}(m-\lambda)^2\frac{\lambda^m}{m!}e^{-\lambda}=e^{-\lambda}\sum_{m=0}^{\infty}(m-\lambda)^2\frac{\lambda^m}{m!}=e^{-\lambda}\sum_{m=0}^{\infty}\left(m^2-2m\lambda+\lambda^2\right)\frac{\lambda^m}{m!}=$$

$$=e^{-\lambda}\left(\sum_{m=0}^{\infty}m^2\frac{\lambda^m}{m!}-\sum_{m=0}^{\infty}2m\lambda\frac{\lambda^m}{m!}+\sum_{m=0}^{\infty}\lambda^2\frac{\lambda^m}{m!}\right)=e^{-\lambda}\left(\sum_{m=1}^{\infty}m^2\frac{\lambda^m}{m!}-\sum_{m=1}^{\infty}2m\lambda\frac{\lambda^m}{m!}+\sum_{m=0}^{\infty}\lambda^2\frac{\lambda^m}{m!}\right)=$$

$$=e^{-\lambda}\left(\sum_{m=1}^{\infty}m\frac{\lambda^m}{(m-1)!}-2\lambda\sum_{m=1}^{\infty}\frac{\lambda^m}{(m-1)!}+\lambda^2\sum_{m=0}^{\infty}\frac{\lambda^m}{m!}\right)=e^{-\lambda}\left(\sum_{m=1}^{\infty}(m-1+1)\frac{\lambda^m}{(m-1)!}-2\lambda\cdot\lambda\sum_{m=1}^{\infty}\frac{\lambda^{m-1}}{(m-1)!}+\lambda^2e^{\lambda}\right)=$$

$$=e^{-\lambda}\left(\sum_{m=1}^{\infty}(m-1)\frac{\lambda^m}{(m-1)!}+\sum_{m=1}^{\infty}\frac{\lambda^m}{(m-1)!}-2\lambda^2e^{\lambda}+\lambda^2e^{\lambda}\right)=e^{-\lambda}\left(\sum_{m=2}^{\infty}(m-1)\frac{\lambda^m}{(m-1)!}+\lambda\sum_{m=1}^{\infty}\frac{\lambda^{m-1}}{(m-1)!}-\lambda^2e^{\lambda}\right)=$$

$$=e^{-\lambda}\left(\sum_{m=2}^{\infty}\frac{\lambda^m}{(m-2)!}+\lambda e^{\lambda}-\lambda^2e^{\lambda}\right)=e^{-\lambda}\left(\lambda^2\sum_{m=2}^{\infty}\frac{\lambda^{m-2}}{(m-2)!}+\lambda e^{\lambda}-\lambda^2e^{\lambda}\right)=e^{-\lambda}\left(\lambda^2e^{\lambda}+\lambda e^{\lambda}-\lambda^2e^{\lambda}\right)=e^{-\lambda}\lambda e^{\lambda}=\lambda$$

Соответственно, среднеквадратическое отклонение будет равно $\sigma(X)=\sqrt{\lambda}$.

Найдём дисперсию равномерно распределённой на промежутке $[a;b)$ случайной величины (напомним, что $EX=(a+b)/2$):

$$DX=\int_{-\infty}^{+\infty}(x-EX)^2 f(x)dx=\int_{-\infty}^{+\infty}\left(x-\frac{a+b}{2}\right)^2 f(x)dx=\int_{-\infty}^{a}\left(x-\frac{a+b}{2}\right)^2 f(x)dx+$$
$$+\int_{a}^{b}\left(x-\frac{a+b}{2}\right)^2 f(x)dx+\int_{b}^{+\infty}\left(x-\frac{a+b}{2}\right)^2 f(x)dx=\int_{-\infty}^{a}\left(x-\frac{a+b}{2}\right)^2 0dx+\int_{a}^{b}\left(x-\frac{a+b}{2}\right)^2 \frac{1}{b-a}dx+$$
$$+\int_{b}^{+\infty}\left(x-\frac{a+b}{2}\right)^2 0dx=0+\frac{1}{b-a}\int_{a}^{b}\left(x-\frac{a+b}{2}\right)^2 dx+0=\frac{1}{b-a}\cdot\frac{1}{3}\cdot\left(x-\frac{a+b}{2}\right)^3\Bigg|_a^b=$$
$$=\frac{1}{b-a}\cdot\frac{1}{3}\cdot\left(\left(b-\frac{a+b}{2}\right)^3-\left(a-\frac{a+b}{2}\right)^3\right)=\frac{1}{b-a}\cdot\frac{1}{3}\cdot\left(\left(\frac{2b-a-b}{2}\right)^3-\left(\frac{2a-a-b}{2}\right)^3\right)=$$
$$=\frac{1}{b-a}\cdot\frac{1}{3}\cdot\left(\frac{(b-a)^3}{8}-\frac{(a-b)^3}{8}\right)=\frac{1}{b-a}\cdot\frac{1}{3}\cdot\frac{2(b-a)^3}{8}=\frac{(b-a)^2}{12}.$$

Среднеквадратическое отклонение окажется равным $\sigma(X)=\dfrac{b-a}{2\sqrt{3}}$. Действительно, чем короче промежуток $[a;b)$, тем меньше разность $(b-a)$, и тем меньше дисперсия и среднеквадратическое отклонение, то есть разбросанность случайной величины. Полезно заметить, что дисперсия и среднеквадратическое отклонение зависят только от длины промежутка и не зависят от положения этого промежутка на числовой оси.

Найдём дисперсию нормально распределённой случайной величины. Напомним, что $EX=a$. Тогда $DX=\int_{-\infty}^{+\infty}(x-EX)^2 f(x)dx=\int_{-\infty}^{+\infty}(x-a)^2\cdot\frac{1}{\sigma}\cdot\varphi\left(\frac{x-a}{\sigma}\right)dx$. Сделаем уже знакомую замену переменой: $t=\dfrac{x-a}{\sigma}\Leftrightarrow x-a=t\sigma\Leftrightarrow x=a+\sigma\cdot t\Rightarrow dx=\sigma\cdot dt$, тогда при $x\to-\infty\Leftrightarrow$ $\Leftrightarrow t\to-\infty$, при $x\to+\infty\Leftrightarrow t\to+\infty$. Поэтому: $DX=\int_{-\infty}^{+\infty}(t\sigma)^2\cdot\frac{1}{\sigma}\cdot\varphi(t)\cdot\sigma\cdot dt=\sigma^2\int_{-\infty}^{+\infty}t^2\varphi(t)dt$. Этот интеграл возьмём по частям: $u=t\Rightarrow du=dt$; $dv=t\varphi(t)dt\Rightarrow v=\int t\varphi(t)dt=\int t\cdot\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{t^2}{2}}dt$. В этом интеграле тоже сделаем замену переменной $y=-t^2/2\Rightarrow dy=-t\cdot dt$. И тогда $v=-\frac{1}{\sqrt{2\pi}}\int e^y dy=-\frac{e^y}{\sqrt{2\pi}}=-\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{t^2}{2}}$. Теперь, учитывая чётность интегрируемой функции, симметричность промежутка интегрирования относительно нуля, условие нормированности и используя правило Лопиталя, получим

$$DX=\sigma^2\left(2\lim_{b\to+\infty}\left(-\frac{t}{\sqrt{2\pi}}\cdot e^{-\frac{t^2}{2}}\right)\Bigg|_0^b-\int_{-\infty}^{+\infty}-\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{t^2}{2}}dt\right)=\sigma^2\left(-\frac{2}{\sqrt{2\pi}}\lim_{b\to+\infty}\left(\frac{b}{e^{\frac{b^2}{2}}}+0\right)+\int_{-\infty}^{+\infty}\varphi(t)dt\right)=$$
$$=\sigma^2\left(-\sqrt{\frac{2}{\pi}}\lim_{b\to+\infty}\left(\frac{1}{be^{\frac{b^2}{2}}}\right)+1\right)=\sigma^2\left(-\sqrt{\frac{2}{\pi}}\cdot 0+1\right)=\sigma^2.$$

А среднеквадратическое отклонение, соответственно, равно $\sigma(X)=\sigma$. Дисперсия и среднеквадратическое отклонение, как и для равномерно распределённой случайной величины, оказались независящими от сдвига случайной величины по оси координат.

Найдём дисперсию экспоненциального распределения с параметром λ. Напомним, что $EX=1/\lambda$. Тогда $DX=\int_{-\infty}^{+\infty}(x-EX)^2 f(x)dx=\int_{-\infty}^{0}\left(x-\frac{1}{\lambda}\right)^2 0dx+\int_{0}^{+\infty}\left(x-\frac{1}{\lambda}\right)^2\lambda e^{-\lambda x}dx=$ $=0+\lim\limits_{b\to+\infty}\int_0^b\left(x-\frac{1}{\lambda}\right)^2\lambda e^{-\lambda x}dx$. Интеграл возьмём по частям. Обозначим $u=(x-1/\lambda)^2\Rightarrow$ $\Rightarrow du=2(x-1/\lambda)dx$; $dv=\lambda e^{-\lambda x}dx\Rightarrow v=-e^{-\lambda x}$, тогда $DX=\lim\limits_{b\to+\infty}\left(-\frac{\left(x-\frac{1}{\lambda}\right)^2}{e^{\lambda x}}\Bigg|_0^b+\int_0^b 2\left(x-\frac{1}{\lambda}\right)e^{-\lambda x}dx\right)$.

Ещё раз применим правило интегрирования по частям: $u=2(x-1/\lambda)\Rightarrow du=2dx$; $dv=e^{-\lambda x}dx\Rightarrow v=-(1/\lambda)\cdot e^{-\lambda x}$. Теперь

$$DX=\lim_{b\to+\infty}\left(-\frac{\left(b-\frac{1}{\lambda}\right)^2}{e^{\lambda b}}+\left(-\frac{1}{\lambda}\right)^2-2\left(x-\frac{1}{\lambda}\right)\frac{1}{\lambda}e^{-\lambda x}\Bigg|_0^b+\int_0^b\frac{2}{\lambda}e^{-\lambda x}dx\right)=$$

$$=-\left(\lim_{b\to+\infty}\frac{\left(b-\frac{1}{\lambda}\right)^2}{e^{\lambda b}}\right)+\left(\frac{1}{\lambda}\right)^2-2\lim_{b\to+\infty}\left(b-\frac{1}{\lambda}\right)\frac{1}{\lambda}e^{-\lambda b}+2\left(-\frac{1}{\lambda}\right)\frac{1}{\lambda}-\lim_{b\to+\infty}\left(\frac{2}{\lambda}\cdot\frac{1}{\lambda}\cdot e^{-\lambda x}\right)\Bigg|_0^b=$$

$$=-\left(\lim_{b\to+\infty}\frac{2\left(b-\frac{1}{\lambda}\right)}{\lambda e^{\lambda b}}\right)+\frac{1}{\lambda^2}-\frac{2}{\lambda}\lim_{b\to+\infty}\frac{\left(b-\frac{1}{\lambda}\right)}{e^{\lambda b}}-\frac{2}{\lambda^2}-\lim_{b\to+\infty}\left(\frac{2}{\lambda^2}\cdot e^{-\lambda b}-\frac{2}{\lambda^2}\right)=$$

$=-\left(\lim\limits_{b\to+\infty}\frac{2}{\lambda^2 e^{\lambda b}}\right)-\frac{1}{\lambda^2}-\frac{2}{\lambda}\lim\limits_{b\to+\infty}\frac{1}{\lambda e^{\lambda b}}-\left(0-\frac{2}{\lambda^2}\right)=0-\frac{1}{\lambda^2}-0+\frac{2}{\lambda^2}=\frac{1}{\lambda^2}$. Среднеквадратическое отклонение окажется равным $\sigma(X)=\sqrt{DX}=\sqrt{\frac{1}{\lambda^2}}=\frac{1}{\lambda}$.

Свойства дисперсии

1. Неотрицательность. Дисперсия $DX=\int_{-\infty}^{+\infty}(x-EX)^2 f(x)dx$ равна интегралу от неотрицательной функции, и следовательно $DX\geq 0$. Можно также сослаться на то, что дисперсия это математическое ожидание от неотрицательной случайной величины (от квадрата). Естественно, что и $\sigma(X)\geq 0$.
2. Дисперсия константы равна нулю. Построим «таблицу» распределения случайной величины $Y=(X-EX)^2$, если $X=C$. Учитывая, что $EX=C$:
3.

x_i:	C
x_i-EX	0
$y_i=(x_i-EX)^2$	0
p_i:	1

Тогда $DC=E0=0$. Естественно, что и $\sigma(X)=0$.

4. Если дисперсия случайной величины равна нулю, то эта случайная величина имеет вырожденное распределение. Действительно, $DX=0 \Leftrightarrow \sum_i (x_i - EX)^2 p_i = 0$, причём, как уже выяснялось, можно считать, что $p_i>0$ строго. А поскольку $(x_i-EX)^2\geq 0$, как квадрат, то $(x_i-EX)^2=0\Leftrightarrow x_i-EX=0\Leftrightarrow x_i=EX$. То есть, возможное значение x_i у случайной величины только одно: $x_1=EX$. Это и означает, что случайная величина X вырожденная.
5. Сомножитель-константа выносится за знак дисперсии <u>в квадрате</u>. Действительно, $D(aX)=E((aX-E(aX))^2)=E((aX-aEX)^2)=E((a(X-EX))^2)=E(a^2(X-EX)^2)=a^2E((X-EX)^2)=$ $=a^2DX$. Соответственно, $\sigma(aX)=\sqrt{D(aX)}=\sqrt{a^2DX}=\sqrt{a^2}\cdot\sqrt{DX}=|a|\cdot\sigma(X)$.
6. Дисперсия суммы <u>независимых</u> случайных величин, определённых на одном вероятностном пространстве равна сумме их дисперсий. Действительно, $D(X+Y)=E(((X+Y)-E(X+Y))^2)=E((X+Y-(EX+EY))^2)=E(((X-EX)+(Y-EY))^2)=$ $E((X-EX)^2+2(X-EX)(Y-EY)+(Y-EY)^2)=E((X-EX)^2)+2E((X-EX)(Y-EY))+E((Y-EY)^2)=$ $=DX+2E(XY-(EX)\cdot Y-X\cdot EY+EX\cdot EY)+DY=$ $=DX+2(E(XY)-E((EX)\cdot Y)-E(X\cdot EY)+E(EX\cdot EY))+DY=$ $=DX+2(E(XY)-(EX)\cdot EY-EX\cdot EY+EX\cdot EY)+DY=DX+2(E(XY)-EX\cdot EY)+DY$. Приведённые выкладки верны и в случае отсутствия независимости случайных величин X и Y. А если они независимые, то математическое ожидание их произведений будет равным произведению их математических ожиданий: $E(XY)=EX\cdot EY$. Поэтому $D(X+Y)=DX+2\cdot 0+DY=DX+DY$.
7. Дисперсия разности <u>независимых</u> случайных величин, определённых на одном вероятностном пространстве равна <u>сумме</u> (также, как и дисперсия суммы!) их дисперсий. Действительно, поскольку из независимости случайных величин X и Y следует независимость случайных величин X и $Z=-Y$, то $D(X-Y)=D(X+Z)=DX+DZ=$ $=DX+D(-Y)=DX+(-1)^2\cdot DY=DX+1\cdot DY=DX+DY$.
7. Дисперсия не изменяется при сдвиге на постоянное слагаемое, то есть $D(X+C)=DX$. Действительно, поскольку вырожденная случайная величина не зависима со случайной величиной X, то по предыдущему свойству $D(X+C)=DX+DC=DX+0=DX$.
8. Дисперсия от линейной комбинации независимых случайных величин. $D(aX\pm bY\pm C)=$ $=D(aX)+D(bY)+DC=a^2DX+b^2DY+0=a^2DX+b^2DY$.
9. Дисперсия случайной величины равна математическому ожиданию квадрата минус квадрат математического ожидания. Действительно: $DX=E((X-EX)^2)=$ $=E(X^2-2X\cdot EX+(EX)^2)=E(X^2)-2EX\cdot EX+(EX)^2=E(X^2)-2(EX)^2+(EX)^2=E(X^2)-(EX)^2$. Эта формула удобна для теоретических исследований, но я не рекомендовал бы её для нахождения дисперсии на практике. Дело в том, что обычно математическое ожидание квадрата случайной величины (в относительном смысле) не сильно отличается от квадрата её математического ожидания, и при вычитании одной величины из другой может произойти потеря значащих цифр.
10. Дисперсия произведения <u>независимых центрированных</u> случайных величин, определённых на одном вероятностном пространстве равна произведению их дисперсий. Уточним сначала, что случайная величина называется *центрированной*, если её математическое ожидание равно нулю. Найдём $D(XY)=$ $=E((XY)^2)-(E(XY))^2=E(X^2\cdot Y^2)-(EX\cdot EY)^2=E(X^2\cdot Y^2)-(0\cdot 0)^2=E(X^2\cdot Y^2)$. Поскольку из независимости случайных величин следует и независимость функций от них, то квадраты X^2 и Y^2 тоже независимые случайные величины. Следовательно, $D(XY)=E(X^2)\cdot E(Y^2)=(E(X^2)-0)\cdot(E(Y^2)-0)=(E(X^2)-(EX)^2)\cdot(E(Y^2)-(EY)^2)=DX\cdot DY$.

11. Дисперсия суммы n попарно независимых (и тем более взаимно независимых) случайных величин равна сумме их дисперсий. По девятому свойству $D\left(\sum_{i=1}^{n} X_i\right) =$
$$= E\left(\left(\sum_{i=1}^{n} X_i\right)^2\right) - \left(E\left(\sum_{i=1}^{n} X_i\right)\right)^2 = E\left(\sum_{i=1}^{n}\sum_{j=1}^{n} X_i X_j\right) - \left(\sum_{i=1}^{n} EX_i\right)^2 =$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{n} E\left(X_i X_j\right) - \sum_{i=1}^{n}\sum_{j=1}^{n}\left(EX_i \cdot EX_j\right) = \sum_{i=1}^{n} E\left((X_i)^2\right) + \sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n}\left(EX_i \cdot EX_j\right) - \sum_{i=1}^{n}\sum_{j=1}^{n}\left(EX_i \cdot EX_j\right) =$$
$$= \sum_{i=1}^{n} E\left((X_i)^2\right) - \sum_{i=1}^{n}\sum_{j=1}^{n}\left(\delta_{ij} \cdot EX_i \cdot EX_j\right) = \sum_{i=1}^{n} E\left((X_i)^2\right) - \sum_{i=1}^{n}\left((EX_i)^2\right) =$$
$$= \sum_{i=1}^{n}\left(E\left((X_i)^2\right) - (EX_i)^2\right) = \sum_{i=1}^{n}\left(DX_i\right).$$
12. Найдём дисперсию среднего арифметического n попарно независимых, имеющих одинаковую дисперсию σ^2 (например, одинаково распределённых) случайных величин. $D\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} D\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} DX_i = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{1}{n^2} n\sigma^2 = \frac{1}{n}\sigma^2$. Как показали расчёты, с ростом числа слагаемых в среднем арифметическом дисперсия уменьшается. Это свойство используется для увеличения точности измерений с помощью увеличения числа экспериментов.
13. Экстремальное свойство дисперсии. Докажем, что выражение $E(X\text{-}a)^2$ достигает своего наименьшего значения при a=EX. Для этого докажем неотрицательность разности: $E(X\text{-}a)^2\text{-}E(X\text{-}EX)^2= E((X\text{-}a)^2\text{-}(X\text{-}EX)^2)=E(X^2\text{-}2aX+a^2\text{-}X^2+2X\cdot \text{E}X\text{-}(EX)^2)=$ $=E(\text{-}2aX+a^2+2X\cdot EX\text{-}(EX)^2)$ $=\text{-}2aEX+E(a^2)+2EX\cdot EX\text{-}E(EX)^2)=\text{-}2aEX+a^2+2(EX)^2\text{-}(EX)^2=$ $=\text{-}2aEX+a^2+(EX)^2=(EX\text{-}a)^2\geq 0$. Поскольку равенство нулю достигается при $a=EX$, то утверждение доказано. Это свойство показывает, что дисперсия характеризует разброс случайной величины вокруг именно математического ожидания, а не чего-то другого.

Дисперсия биномиального распределения

Пусть X_i случайная величина, имеющая распределение Бернулли с параметром p это число успехов в i-ом испытании. Тогда случайная величина Y, имеющая биномиальное распределение с параметрами n и p равна $Y = \sum_{i=1}^{n} X_i$. Причём, поскольку по определению биномиального распределения, случайные величины X_i взаимно независимые, то $DY = D\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} DX_i = \sum_{i=1}^{n} p(1-p) = np(1-p)$.

Правило трёх сигм

Выше упоминалось, что в природе многие случайные величины имеют распределения незначительно отличающиеся от нормального, но при этом их возможные значения заполняют сплошь некоторый ограниченный промежуток, а не всю числовую прямую. Опытным путём было установлено, что этот промежуток незначительно отличается от промежутка $[EX-3\cdot\sigma(X); EX+3\cdot\sigma(X)]$. В этом и состоит *правило трёх сигм*: на практике случайные величины отклоняются от своих математических ожиданий не более, чем на три среднеквадратических отклонения.

Если предположить, что случайная величина X имеет именно нормальное распределение (что практически неосуществимо) с параметрами a и σ, то вероятность того, что X примет значение из этого интервала равна:

$$P(a-3\sigma \le X \le a+3\sigma) = F(a+3\sigma) - F(a-3\sigma) = \Phi\left(\frac{a+3\sigma-a}{\sigma}\right) - \Phi\left(\frac{a-3\sigma-a}{\sigma}\right) =$$

$$= \Phi\left(\frac{3\sigma}{\sigma}\right) - \Phi\left(\frac{-3\sigma}{\sigma}\right) = \Phi(3) - \Phi(-3) = \Phi(3) - \left(1-\Phi(3)\right) = 2\cdot\Phi(3) - 1 \approx 2\cdot 0{,}99865 - 1 = 0{,}9973\,.$$

А вероятность противоположного события $P(|X-a|>3\cdot\sigma)=1-P(a-3\cdot\sigma\le X\le a+3\cdot\sigma)\approx$ $\approx 1-0{,}0073=0{,}0027\approx 1/370$. Поскольку на практике такое событие никогда не происходит, то вероятностью 0,27% принято пренебрегать. События, вероятности которых меньше 0,27% называются *практически невозможными*. Соответственно, события противоположные им, вероятности которых больше, чем 99,73% называются *практически достоверными*. Разумеется вероятность 0,27%, хотя и имеет практическое обоснование, весьма условна и в отдельных случаях может быть как понижена так и повышена. Так, например, если речь идёт о жизни и смерти, то вероятность 1/370 не кажется маленькой, а если решается задача о прогнозе погоды, то вероятность ошибки 1/20 вполне приемлема.

Моменты

Кроме математического ожидания и дисперсии существуют другие численные характеристики случайных величин. Наиболее важными из остальных численных характеристик являются *центральные моменты*. Центральным моментом порядка k случайной величины X называется $\mu_k(X)=E((X-EX)^k)$. Найдём несколько первых центральных моментов:

$\mu_0(X)=E((X-EX)^0)=E1=1$;

$\mu_1(X)=E((X-EX)^1)=E(X-EX)=EX-EX=0$;

$\mu_2(X)=E((X-EX)^2)=DX=(\sigma(X))^2$.

Кроме центральных моментов существуют ещё и начальные. *Начальным моментом* порядка k случайной величины X называется $\alpha_k(X)=E(X^k)$. В частности:

$\alpha_0(X)=E(X^0)=E1=1$;

$\alpha_1(X)=E(X^1)=EX$;

$\alpha_2(X)=E(X^2)$.

С помощью моментов можно записать тождество $DX=E(X^2)-(EX)^2 \Leftrightarrow$ $\Leftrightarrow \mu_2(X)=\alpha_2(X)-(\alpha_1(X))^2$.

Если к названию момента добавить слово «абсолютный», то это будет означать, что математическое ожидание вычисляется от модуля соответствующей случайной величины. Так абсолютным центральным моментом порядка k случайной величины X называется $\nu_k(X)=E|(X-EX)^k|$, а абсолютным начальным моментом порядка k случайной величины X называется $\beta_k(X)=E|X^k|$. Первый абсолютный центральный момент $\nu_1(X)=E|X-EX|$ называется *средним арифметическим отклонением*. При доказательстве существования математического ожидания нормального распределения, фактически было найдено среднее арифметическое отклонение для нормально распределённой случайной величины с математическим ожиданием $EX=a$ и среднеквадратическим отклонением $\sigma(X)=\sigma$. С учётом замены переменной $x=a+\sigma t$ оно было равно:

$$\nu_1(X) = \int_{-\infty}^{+\infty} |x-a|\cdot f(x)dx = \int_{-\infty}^{+\infty} |x-a|\cdot\frac{1}{\sigma}\varphi\left(\frac{x-a}{\sigma}\right)dx = \int_{-\infty}^{+\infty} |\sigma t|\cdot\frac{1}{\sigma}\varphi(t)\sigma dt = 2\sigma\int_{0}^{+\infty} t\varphi(t)dt = \sigma\sqrt{\frac{2}{\pi}}\,.$$

Асимметрия

Третий центральный момент $\mu_3(X)=E((X-EX)^3)$ является простейшим нечётным моментом, не являющимся константой. По знаку центрального момента с нечётным номером большим двух можно определить «скошенность» распределения. Естественно, что её определяют по третьему центральному моменту. Под этим подразумевается, что при $\mu_3(X)>0$ правая половина графика плотности вероятности случайной величины X более пологая, чем левая, а при $\mu_3(X)<0$ левая половина более пологая, чем правая. При симметричном относительно некоторой точки распределении $\mu_3(X)=0$. Интересно, что при $\mu_3(X)=0$ распределение не обязательно симметричное. Заметим, что ничто не мешает подобным образом исследовать скошенности дискретных и смешанных случайных величин.

Третий момент имеет размерность куба случайной величины. А при сравнении друг с другом разных случайных величин удобно использовать безразмерные характеристики. Поэтому для сравнения скошенностей используют *асимметрию*: $S(X)=\dfrac{\mu_3(X)}{(\sigma(X))^3}$.

В качестве примера найдём асимметрию экспоненциального распределения с параметром λ. Напомним, что $EX=1/\lambda$, $DX=\lim\limits_{b\to+\infty}\int\limits_0^b\left(x-\dfrac{1}{\lambda}\right)^2\lambda e^{-\lambda x}dx=\dfrac{1}{\lambda^2}$. Тогда третий центральный момент $\mu_3(X)=\int\limits_{-\infty}^{+\infty}(x-EX)^3f(x)dx=\int\limits_{-\infty}^{0}\left(x-\dfrac{1}{\lambda}\right)^3 0dx+\int\limits_0^{+\infty}\left(x-\dfrac{1}{\lambda}\right)^3\lambda e^{-\lambda x}dx=0+\lim\limits_{b\to+\infty}\int\limits_0^b\left(x-\dfrac{1}{\lambda}\right)^3\lambda e^{-\lambda x}dx$.

Интеграл возьмём по частям. Обозначим $u=(x-1/\lambda)^3\Rightarrow du=3(x-1/\lambda)^2dx$; $dv=\lambda e^{-\lambda x}dx\Rightarrow v=-e^{-\lambda x}$, тогда $\mu_3(X)=\lim\limits_{b\to+\infty}\left(-\left.\dfrac{\left(x-\dfrac{1}{\lambda}\right)^3}{e^{\lambda x}}\right|_0^b+\int\limits_0^b 3\left(x-\dfrac{1}{\lambda}\right)^2e^{-\lambda x}dx\right)=$

$$=\lim_{b\to+\infty}\left(-\frac{\left(b-\frac{1}{\lambda}\right)^3}{e^{\lambda b}}+\left(-\frac{1}{\lambda}\right)^3\right)+\frac{3}{\lambda}\lim_{b\to+\infty}\int_0^b\left(x-\frac{1}{\lambda}\right)^2\lambda e^{-\lambda x}dx=-\lim_{b\to+\infty}\left(\frac{\left(b-\frac{1}{\lambda}\right)^3}{e^{\lambda b}}\right)-\left(\frac{1}{\lambda}\right)^3+\frac{3}{\lambda}DX=$$

$$=-\left(\lim_{b\to+\infty}\frac{3\left(b-\frac{1}{\lambda}\right)^2}{\lambda e^{\lambda b}}\right)-\frac{1}{\lambda^3}+\frac{3}{\lambda}\cdot\frac{1}{\lambda^2}=-\left(\lim_{b\to+\infty}\frac{6\left(b-\frac{1}{\lambda}\right)}{\lambda^2e^{\lambda b}}\right)-\frac{1}{\lambda^3}+\frac{3}{\lambda^3}=-\left(\lim_{b\to+\infty}\frac{6}{\lambda^3e^{\lambda b}}\right)+\frac{2}{\lambda^3}=$$

$=0+\dfrac{2}{\lambda^3}=\dfrac{2}{\lambda^3}$. Напомним, что среднеквадратическое отклонение экспоненциального распределения равно $\sigma(X)=1/\lambda$. Поэтому асимметрия экспоненциального распределения $S(X)=\dfrac{\mu_3(X)}{\sigma^3(X)}=\dfrac{2/\lambda^3}{(1/\lambda)^3}=\dfrac{2\cdot\lambda^3}{\lambda^3}=2$. Асимметрия положительная, и это арифметически доказывает, что скошенность экспоненциального распределения направлена в правую сторону, что геометрически соответствует тому, что правая половина графика её плотности распределения более пологая, чем левая (см. рис. 27).

Найдём асиметрию дискретной случайной величины, имеющей распределение Бернулли с параметром p. Вспомним её таблицу распределения.

x_i:	0	1
x_i-EX	$-p$	$1-p$
$(x_i-EX)^3$	$-p^3$	$(1-p)^3$
p_i:	$1-p$	p

$\mu_3(X)=-p^3(1-p)+(1-p)^3p=p(1-p)(-p^2+(1-p)^2)=p(1-p)(-p^2+1-2p+p^2)=p(1-p)(1-2p)$.

Вспомним, что $\sigma(X)=\sqrt{p(1-p)}$. Тогда $S(X)=\dfrac{\mu_3(X)}{(\sigma(X))^3}=\dfrac{p(1-p)(1-2p)}{\left(\sqrt{p(1-p)}\right)^3}=\dfrac{1-2p}{\sqrt{p(1-p)}}$.

При $p<½$ асиметрия $S(X)$ отрицательная. Геометрический смысл отрицательности асимметрии в том, что скошенность *многоугольника распределения* (фигуры под ломаной линией, соединяющей точки (x_i, p_i) с соседними номерами) направлена в левую сторону. Действительно при $p<½$ вероятность того, что случайная величина примет левое из возможных значений меньше, чем правое. При $p>½$ скошенность многоугольника распределения направлена в правую сторону.

Эксцесс

Четвёртый центральный момент $\mu_4(X)=E((X-EX)^4)$ служит характеристикой «крутости» (островершинности) графика плотности вероятности случайной величины X. Найдем его для нормального распределения с параметрами a и σ. Напомним, что $EX=a$. Тогда четвёртый центральный момент $\mu_4(X)=\int\limits_{-\infty}^{+\infty}(x-EX)^4 f(x)dx=\int\limits_{-\infty}^{+\infty}(x-a)^4\cdot\frac{1}{\sigma}\cdot\varphi\left(\frac{x-a}{\sigma}\right)dx$. В очередной раз сделаем замену переменой: $t=\dfrac{x-a}{\sigma}\Leftrightarrow x-a=t\sigma\Leftrightarrow x=a+\sigma\cdot t\Rightarrow dx=\sigma\cdot dt$, тогда при $x\to-\infty \Leftrightarrow t\to-\infty$, при $x\to+\infty \Leftrightarrow$ $\Leftrightarrow t\to+\infty$. И поэтому: $\mu_4(X)=\int\limits_{-\infty}^{+\infty}(t\sigma)^4\cdot\frac{1}{\sigma}\cdot\varphi(t)\cdot\sigma\cdot dt=\sigma^4\int\limits_{-\infty}^{+\infty}t^4\varphi(t)dt$. Этот интеграл возьмём по частям: $u=t^3\Rightarrow du=3t^2dt$; $dv=t\varphi(t)dt\Rightarrow v=\int t\varphi(t)dt=\int t\cdot\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{t^2}{2}}dt$. В этом интеграле тоже сделаем замену переменной $y=-t^2/2\Rightarrow dy=-t\cdot dt$. И тогда $v=-\frac{1}{\sqrt{2\pi}}\int e^y dy=-\frac{e^y}{\sqrt{2\pi}}=-\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{t^2}{2}}$. Теперь, учитывая чётность интегрируемой функции, симметричность промежутка интегрирования относительно нуля, дважды используя правило Лопиталя и вспоминая, что для нормального распределения $\sigma^2=DX=\sigma^2\int\limits_{-\infty}^{+\infty}t^2\varphi(t)dt\Rightarrow\int\limits_{-\infty}^{+\infty}t^2\varphi(t)dt=1$, получим $\mu_4(X)=$

$$=\sigma^4\left(2\lim_{b\to+\infty}\left(-\frac{t^3}{\sqrt{2\pi}}\cdot e^{-\frac{t^2}{2}}\right)\Bigg|_0^b-\int\limits_{-\infty}^{+\infty}-\frac{3t^2}{\sqrt{2\pi}}\cdot e^{-\frac{t^2}{2}}dt\right)=\sigma^4\left(-\frac{2}{\sqrt{2\pi}}\lim_{b\to+\infty}\left(\frac{b^3}{e^{\frac{b^2}{2}}}+0\right)+\int\limits_{-\infty}^{+\infty}3t^2\varphi(t)dt\right)=$$

$$=\sigma^4\left(-\sqrt{\frac{2}{\pi}}\lim_{b\to+\infty}\left(\frac{3b^2}{be^{\frac{b^2}{2}}}\right)+3\int\limits_{-\infty}^{+\infty}t^2\varphi(t)dt\right)=\sigma^4\left(-\sqrt{\frac{2}{\pi}}\lim_{b\to+\infty}\left(\frac{3b}{e^{\frac{b^2}{2}}}\right)+3\cdot1\right)=$$

$=\sigma^4\left(-\sqrt{\frac{2}{\pi}}\lim\limits_{b\to+\infty}\left(\frac{3}{be^{\frac{b^2}{2}}}\right)+3\right)=\sigma^4\left(-\sqrt{\frac{2}{\pi}}\cdot3\cdot0+3\right)=3\sigma^4$. Четвёртый момент имеет размерность четвёртой степени случайной величины. Чтобы получить безразмерную характеристику крутости графика плотности вероятностей разделим четвёртый центральный момент на четвёртую степень среднеквадратического отклонения. Напомним, что для нормального распределения $\sigma(X)=\sigma$. Получится, что для нормального распределения $\frac{\mu_4(X)}{(\sigma(X))^4}=\frac{3\sigma^4}{\sigma^4}=3$. Если исходить из того, что нормальное распределение должно быть нормальным во всех отношениях, и крутость графика плотности его распределения должна быть равна нулю, то можно предложить характеристику, называемую *эксцесс*: $\varepsilon(X)=\frac{\mu_4(X)}{(\sigma(X))^4}-3$. Для нормального распределения $\varepsilon(X)=0$.

Найдём эксцесс случайной величины, равномерно распределённой на промежутке $[a;b)$. Напомним, что $EX=\frac{a+b}{2}$. Тогда четвёртый центральный момент

$$\mu_4(X)=\int_{-\infty}^{+\infty}(x-EX)^4 f(x)dx=\int_{-\infty}^{a}\left(x-\frac{a+b}{2}\right)^4 0dx+\int_{a}^{b}\left(x-\frac{a+b}{2}\right)^4\frac{dx}{b-a}+\int_{b}^{+\infty}\left(x-\frac{a+b}{2}\right)^4 0dx=$$

$$=0+\frac{1}{5}\left(x-\frac{a+b}{2}\right)^5\frac{1}{b-a}\Bigg|_a^b+0=\frac{1}{5}\left(\left(b-\frac{a+b}{2}\right)^5-\left(a-\frac{a+b}{2}\right)^5\right)\frac{1}{b-a}=$$

$$=\frac{1}{5}\left(\left(\frac{b-a}{2}\right)^5-\left(\frac{a-b}{2}\right)^5\right)\frac{1}{b-a}=\frac{2}{5}\left(\frac{b-a}{2}\right)^5\frac{1}{b-a}=\frac{(b-a)^4}{5\cdot 2^4}=\frac{(b-a)^4}{80}.$$

Напомним, что для равномерного распределения $\sigma(X)=\frac{b-a}{2\sqrt{3}}$. Тогда $\varepsilon(X)=\frac{\mu_4(X)}{(\sigma(X))^4}-3=$

$$=\frac{(b-a)^4/5\cdot 2^4}{(b-a)^4/9\cdot 2^4}-3=\frac{9}{5}-3=-\frac{6}{5}.$$

Эксцесс равномерного распределения отрицательный, то есть график плотности его распределения имеет меньшую крутизну, чем график плотности нормального распределения.

Найдём эксцесс случайной величины имеющей экспоненциальное распределение с параметром λ. Напомним, что $EX=1/\lambda$, $\mu_3(X)=\lim\limits_{b\to+\infty}\int_0^b\left(x-\frac{1}{\lambda}\right)^3\lambda e^{-\lambda x}dx=\frac{2}{\lambda^3}$. Тогда четвёртый центральный момент

$$\mu_4(X)=\int_{-\infty}^{+\infty}(x-EX)^4 f(x)dx=\int_{-\infty}^{0}\left(x-\frac{1}{\lambda}\right)^4 0dx+\int_{0}^{+\infty}\left(x-\frac{1}{\lambda}\right)^4\lambda e^{-\lambda x}dx=0+\lim_{b\to+\infty}\int_0^b\left(x-\frac{1}{\lambda}\right)^4\lambda e^{-\lambda x}dx.$$

Интеграл возьмём по частям. Обозначим $u=(x-1/\lambda)^4\Rightarrow du=4(x-1/\lambda)^3dx$; $dv=\lambda e^{-\lambda x}dx\Rightarrow v=-e^{-\lambda x}$, тогда $\mu_4(X)=\lim\limits_{b\to+\infty}\left(-\frac{\left(x-\frac{1}{\lambda}\right)^4}{e^{\lambda x}}\Bigg|_0^b+\int_0^b 4\left(x-\frac{1}{\lambda}\right)^3 e^{-\lambda x}dx\right)=$

$$=\lim_{b\to+\infty}\left(-\frac{\left(b-\frac{1}{\lambda}\right)^4}{e^{\lambda b}}+\left(-\frac{1}{\lambda}\right)^4\right)+\frac{4}{\lambda}\lim_{b\to+\infty}\int_0^b\left(x-\frac{1}{\lambda}\right)^3\lambda e^{-\lambda x}dx=-\lim_{b\to+\infty}\left(\frac{\left(b-\frac{1}{\lambda}\right)^4}{e^{\lambda b}}\right)+\left(\frac{1}{\lambda}\right)^4+\frac{4}{\lambda}\mu_3(X)=$$

$$=-\left(\lim_{b\to+\infty}\frac{4\left(b-\frac{1}{\lambda}\right)^3}{\lambda e^{\lambda b}}\right)+\frac{1}{\lambda^4}+\frac{4}{\lambda}\cdot\frac{2}{\lambda^3}=-\left(\lim_{b\to+\infty}\frac{12\left(b-\frac{1}{\lambda}\right)^2}{\lambda^2 e^{\lambda b}}\right)+\frac{1}{\lambda^4}+\frac{8}{\lambda^4}=-\left(\lim_{b\to+\infty}\frac{24\left(b-\frac{1}{\lambda}\right)}{\lambda^3 e^{\lambda b}}\right)+\frac{9}{\lambda^4}=$$

$=-\left(\lim_{b\to+\infty}\frac{24}{\lambda^4 e^{\lambda b}}\right)+\frac{9}{\lambda^4}=0+\frac{9}{\lambda^4}=\frac{9}{\lambda^4}$. Напомним, что среднеквадратическое отклонение экспоненциального распределения равно $\sigma(X)=1/\lambda$. Поэтому эксцесс экспоненциального распределения $\varepsilon(X)=\frac{\mu_4(X)}{\sigma^4(X)}-3=\frac{9/\lambda^4}{(1/\lambda)^4}-3=9-3=6$. Он больше нуля, и это означает, что график плотности экспоненциального имеет большую крутизну, чем график плотности нормального распределения.

Найдём эксцесс дискретной случайной величины, имеющей распределение Бернулли с параметром p. Вспомним её таблицу распределения.

x_i:	0	1
x_i-EX	$-p$	$1-p$
$(x_i-EX)^4$	p^4	$(1-p)^4$
p_i:	$1-p$	p

$\mu_4(X)=p^4(1-p)+(1-p)^4p=p(1-p)(p^3+(1-p)^3)=p(1-p)(p^3+1-3p+3p^2-p^3)=p(1-p)(1-3p+3p^2)$.
Вспомним, что $\sigma(X)=\sqrt{p(1-p)}$. Тогда
$\varepsilon(X)=\frac{\mu_4(X)}{(\sigma(X))^4}-3=\frac{p(1-p)(1-3p+3p^2)}{p^2(1-p)^2}-3=\frac{1-3p+3p^2-3p(1-p)}{p(1-p)}=\frac{1-6p+6p^2}{p(1-p)}$.

Корнями числителя являются числа $p_{1,2}=\frac{1}{2}\pm\frac{1}{2\sqrt{3}}$. Они примерно равны: $p_1\approx 0{,}21$ и $p_2\approx 0{,}79$. Получается, что если вероятность того, что случайная величина примет одно из возможных значений одна пятая или меньше, то эксцесс распределения Бернулли положительный, перекос слишком велик и крутизна больше, чем у нормального распределения.

Мода

Модой случайной величины называется наиболее вероятное её возможное значение по сравнению с соседними возможными значениями. Из-за парадокса нулевой вероятности это определение нельзя формально применить для непрерывных случайных величин. Для непрерывных случайных величин модой считается то возможное значение, при котором плотность вероятности достигает (локального) максимума. Например, мода нормального распределения совпадает с математическим ожиданием, а мода экспоненциального распределения равна нулю.

Локальных максимумов плотности вероятности может быть два или более. В этом случае случайная величина называется *бимодальной* или *полимодальной*.

Пусть существуют две нормально распределённые случайные величины X и Y и математическими ожиданиями $EX=a$ и $EY=b$ и их дисперсии равны друг другу: $DX=DY=\sigma^2$. Рассмотрим случайную величину Z, построенную с помощью двухступенчатого эксперимента. Подбрасывается симметричная монета. Если выпадает герб, то случайная величина Z принимается равной случайной величине X, а если выпадает решка, то случайной величине Y. Найдём функцию распределения случайной величины Z. По формуле полной вероятности $F(x)=P(Z<x)=$ $=\frac{1}{2}\cdot P(X<x)+\frac{1}{2}\cdot P(Y<x)=\frac{1}{2}\cdot\Phi\left(\frac{x-a}{\sigma}\right)+\frac{1}{2}\cdot\Phi\left(\frac{x-b}{\sigma}\right)$. Плотность вероятности случайной величины Z найдем дифференцированием функции распределения: $f(x)=(F(x))'=\left(\frac{1}{2}\cdot\Phi\left(\frac{x-a}{\sigma}\right)+\frac{1}{2}\cdot\Phi\left(\frac{x-b}{\sigma}\right)\right)'=\frac{1}{2\sigma}\varphi\left(\frac{x-a}{\sigma}\right)+\frac{1}{2\sigma}\varphi\left(\frac{x-b}{\sigma}\right)$. Чтобы выяснить, есть ли у этой функции максимумы найдём её производную: $f'(x)=$ $=\left(\frac{1}{2\sigma}\varphi\left(\frac{x-a}{\sigma}\right)+\frac{1}{2\sigma}\varphi\left(\frac{x-b}{\sigma}\right)\right)'=-\frac{1}{2\sigma}\left(\frac{x-a}{\sigma}\right)\cdot\varphi\left(\frac{x-a}{\sigma}\right)\cdot\frac{1}{\sigma}-\frac{1}{2\sigma}\left(\frac{x-b}{\sigma}\right)\cdot\varphi\left(\frac{x-b}{\sigma}\right)\cdot\frac{1}{\sigma}$. Перед нахождением этой производной полезно сначала вывести формулу для дифференцирования функции $(\varphi(x))'=\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\right)'=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\cdot\left(\frac{-2x}{2}\right)=-x\varphi(x)$. В силу симметрии, производная $f'(x)$ имеет корень $x_0=\frac{a+b}{2}$. Покажем это: $f'\left(\frac{a+b}{2}\right)=$

$$=-\frac{1}{2\sigma}\left(\frac{\frac{a+b}{2}-a}{\sigma}\right)\cdot\varphi\left(\frac{\frac{a+b}{2}-a}{\sigma}\right)\cdot\frac{1}{\sigma}-\frac{1}{2\sigma}\left(\frac{\frac{a+b}{2}-b}{\sigma}\right)\cdot\varphi\left(\frac{\frac{a+b}{2}-b}{\sigma}\right)\cdot\frac{1}{\sigma}=$$

$$=-\frac{1}{2\sigma^3}\left(\frac{b-a}{2}\right)\cdot\varphi\left(\frac{b-a}{2\sigma}\right)-\frac{1}{2\sigma^3}\left(\frac{a-b}{2}\right)\cdot\varphi\left(\frac{a-b}{2\sigma}\right)=\frac{1}{2\sigma^3}\left(\frac{b-a}{2}\right)\cdot\left(-\varphi\left(\frac{b-a}{2\sigma}\right)+\varphi\left(\frac{a-b}{2\sigma}\right)\right)=$$

$=\frac{1}{2\sigma^3}\left(\frac{b-a}{2}\right)\cdot 0=0$. При этом использовано, что функция $\varphi(x)$ чётная. Чтобы выяснить тип экстремума в точке x_0 найдем в ней знак второй производной. Для этого сначала второй раз продифференцируем плотность вероятности: $f''(x)=$

$$=-\frac{1}{2\sigma^3}\left((x-a)\cdot\varphi\left(\frac{x-a}{\sigma}\right)+(x-b)\cdot\varphi\left(\frac{x-b}{\sigma}\right)\right)'=$$

$$=-\frac{1}{2\sigma^3}\left(\varphi\left(\frac{x-a}{\sigma}\right)-(x-a)\left(\frac{x-a}{\sigma}\right)\varphi\left(\frac{x-a}{\sigma}\right)\cdot\frac{1}{\sigma}+\varphi\left(\frac{x-b}{\sigma}\right)-(x-b)\left(\frac{x-b}{\sigma}\right)\varphi\left(\frac{x-b}{\sigma}\right)\cdot\frac{1}{\sigma}\right)=$$

$=-\frac{1}{2\sigma^3}\left(\left(1-\left(\frac{x-a}{\sigma}\right)^2\right)\cdot\varphi\left(\frac{x-a}{\sigma}\right)+\left(1-\left(\frac{x-b}{\sigma}\right)^2\right)\cdot\varphi\left(\frac{x-b}{\sigma}\right)\right)$. А теперь подставим точку $x_0=\frac{a+b}{2}$ в полученную вторую производную: $f''\left(\frac{a+b}{2}\right)=$

$$=-\frac{1}{2\sigma^3}\left(\left(1-\left(\frac{\frac{a+b}{2}-a}{\sigma}\right)^2\right)\cdot\varphi\left(\frac{\frac{a+b}{2}-a}{\sigma}\right)+\left(1-\left(\frac{\frac{a+b}{2}-b}{\sigma}\right)^2\right)\cdot\varphi\left(\frac{\frac{a+b}{2}-b}{\sigma}\right)\right)=$$

$$= -\frac{1}{2\sigma^3}\left(\left(1-\left(\frac{b-a}{2\sigma}\right)^2\right)\cdot\varphi\left(\frac{b-a}{2\sigma}\right)+\left(1-\left(\frac{a-b}{2\sigma}\right)^2\right)\cdot\varphi\left(\frac{a-b}{2\sigma}\right)\right) = -\frac{1}{\sigma^3}\left(1-\left(\frac{b-a}{2\sigma}\right)^2\right)\cdot\varphi\left(\frac{b-a}{2\sigma}\right).$$

Если это выражение положительное, то согласно достаточному признаку экстремума по второй производной плотность вероятности имеет в точке x_0 минимум. Это произойдёт при $1-\left(\frac{b-a}{2\sigma}\right)^2 < 0 \Leftrightarrow \left(\frac{b-a}{2\sigma}\right)^2 > 1$. Учитывая положительность дроби: $\frac{b-a}{2\sigma} > 1 \Leftrightarrow b-a > 2\sigma$. Получается, что если математические ожидания (или моды, что для нормального распределения то же самое) случайных величин X и Y находятся достаточно далеко друг от друга, то в точке симметрии x_0 плотность вероятности имеет минимум, что при положительности этой функции и того, что $\lim_{x\to\infty} f(x) = 0$ обеспечивает существование максимумов слева и справа от минимума. На рисунке 32 показаны графики плотностей вероятностей случайной величины Z. Графики, обозначенные цифрами, построены при: 1: a=0, b=8; 2: a=1, b=7; 3: a=2, b=6; 4: a=3, b=5; 5: a=b=4. Для всех кривых σ=2. В пятом случае, это обычное нормальное распределение.

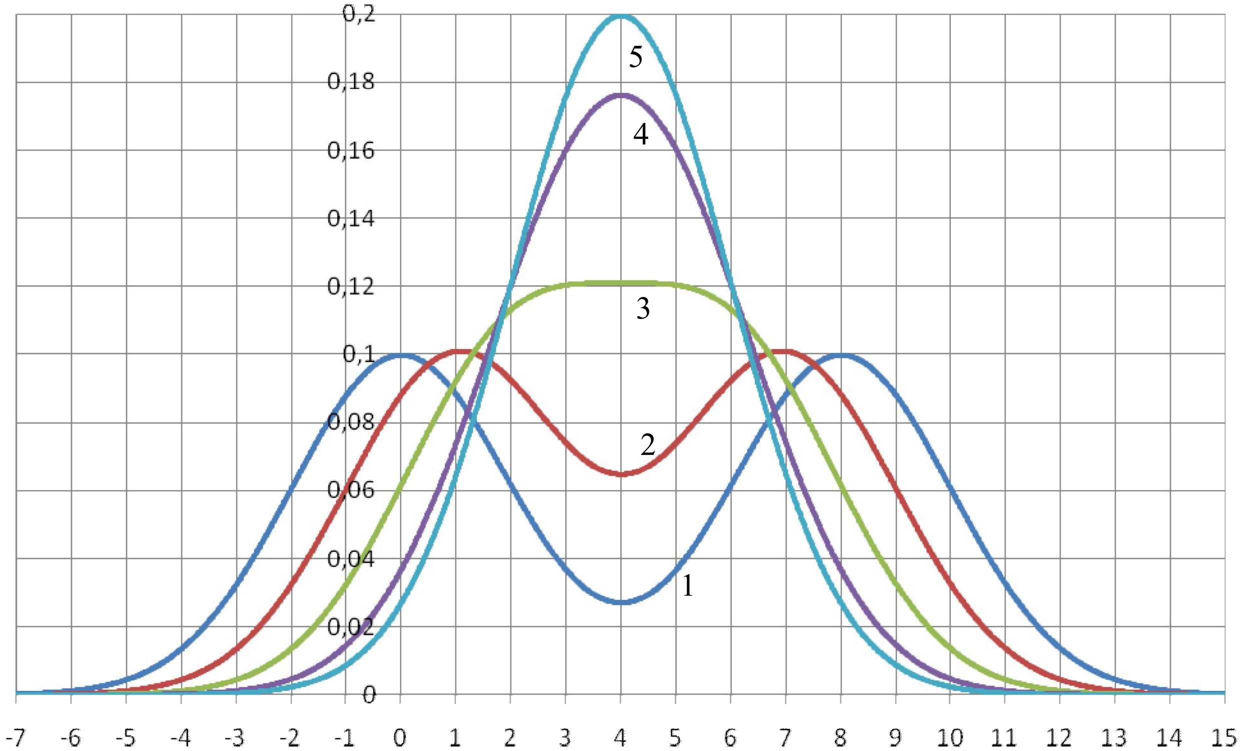

Рис. 32. Возможность образования бимодального и модального распределений в двухступенчатом эксперименте в зависимости от отношения разности математических ожиданий и удвоенного среднеквадратического отклонения.

Медиана

Медианой случайной величины с функцией распределения $F(x)$ называется такое число m, что $F(m)\leq½$, и $\lim_{x\to m+0} F(x) \geq ½$. Заметим, что для дискретных случайных величин медиана определена неоднозначно и может быть произвольным числом некоторого промежутка, а для непрерывных случайных величин единственна, причём медианой является такое возможное значение непрерывной случайной величины, для которого одинаково вероятно, окажется ли случайная величина больше или меньше этого значения.

Для непрерывной случайной величины X будет справедливо равенство: $P(X<m)=P(X>m)\Leftrightarrow F(m)=1-P(X\leq m)\Leftrightarrow F(m)=1-P(X<m)-P(X=m)\Leftrightarrow F(m)=1-F(m)-0\Leftrightarrow$ $\Leftrightarrow 2F(m)=1\Leftrightarrow F(m)=½$.

Найдём медиану экпоненциального распределения с параметром λ. Вспомним, что при положительных значениях аргумента функция распределения экспоненциального распределения $F(m)=1-\frac{1}{e^{\lambda m}}$. Тогда $1-\frac{1}{e^{\lambda m}}=\frac{1}{2}\Leftrightarrow\frac{1}{e^{\lambda m}}=\frac{1}{2}\Leftrightarrow e^{\lambda m}=2\Leftrightarrow\lambda m=\ln 2\Leftrightarrow m=\frac{\ln 2}{\lambda}\approx\frac{0{,}693}{\lambda}$.

Ковариация

Ковариацией (или *коэффициентом ковариации*) двух случайных величин, имеющих математические ожидания и определённых на одном вероятностном пространстве называется величина $\operatorname{cov}(X,Y)=E((X-EX)(Y-EY))$.

Заметим, что $\operatorname{cov}(X,Y)=E((X-EX)(Y-EY))=E(XY-Y\cdot EX-X\cdot EY+EX\cdot EY)=$ $=E(XY)-E(Y\cdot EX)-E(X\cdot EY)+E(EX\cdot EY)=E(XY)-EX\cdot EY-EY\cdot EX+EX\cdot EY=E(XY)-EX\cdot EY$.

Если случайные величины X и Y независимые, то математическое ожидание их произведения равно произведению их математических ожиданий. Поэтому при независимости случайных величин их ковариация равна нулю.

Обратное не всегда верно. То есть, если ковариация равна нулю, то это ещё не означает, что случайные величины обязательно независимые. Для доказательства этого утверждения достаточно привести хотя бы один пример зависимых случайных величин, ковариация которых будет равна нулю. Пусть X дискретная случайная величина, заданная таблицей распределения:

x_i:	−1	0	1
p_i:	1/3	1/3	1/3

И пусть случайная величина Y определяется соотношением $Y=X^2$. В таком случае между случайными величинами есть не просто зависимость, а величина Y однозначно определяется по величине X. Совместная таблица распределения будет иметь вид:

x_i:	−1	0	1
y_i:	1	0	1
$x_i\, y_i$:	−1	0	1
p_i:	1/3	1/3	1/3

Очевидно, что $EX=E(XY)=-1\cdot(1/3)+0\cdot(1/3)+1\cdot(1/3)=0$; $EY=1\cdot(1/3)+0\cdot(1/3)+1\cdot(1/3)=2/3$; $\operatorname{cov}(X,Y)=E(XY)-EX\cdot EY=0-0\cdot(2/3)=0$.

Итак, ковариация иногда позволяет получить некоторую информацию о зависимости случайных величин.

Интересно, что используя термин ковариации первую часть (верную и при зависимых случайных величинах X и Y) пятого свойства дисперсии можно записать так: $D(X+Y)=DX+2\cdot\operatorname{cov}(X,Y)+DY$.

Корреляция

Корреляцией (или *коэффициентом корреляции*) двух случайных величин, определённых на одном вероятностном пространстве и имеющих математические ожидания и ненулевые дисперсии (или среднеквадратические отклонения) называется величина $\rho(X,Y)=\frac{\operatorname{cov}(X,Y)}{\sigma(X)\cdot\sigma(Y)}$. Корреляция, как и ковариация, также иногда позволяет получить некоторую информацию о зависимости случайных величин, но корреляция несколько удобнее, поскольку является безразмерной величиной. Очевидно, что найденные свойства ковариации полностью справедливы и для корреляции: если случайные величины независимые, то их корреляция равна нулю, причём обратное не всегда верно, то есть если корреляция равна нулю, то это не означает, что случайные величины независимые. Случайные величины, корреляция которых отлична от нуля называются *коррелирующими*. И напротив, если корреляция равна нулю, то *некоррелирующими*. Заметим, что словом «коррелировать» нельзя заменять слово «зависеть».

Рассмотрим другие свойства корреляции.

1. Модуль корреляции не превосходит единицы.

Для доказательства этого рассмотрим математические ожидания двух неотрицательных случайных величин $Z_{1,2}^2=\left(\frac{X-EX}{\sigma(X)}\pm\frac{Y-EY}{\sigma(Y)}\right)^2$. По одному из свойств математического ожидания они также оба неотрицательные: $0\le E\left(Z_{1,2}^2\right)=E\left(\left(\frac{X-EX}{\sigma(X)}\pm\frac{Y-EY}{\sigma(Y)}\right)^2\right)=$

$$=E\left(\frac{(X-EX)^2}{(\sigma(X))^2}\pm2\frac{(X-EX)(Y-EY)}{\sigma(X)\cdot\sigma(Y)}+\frac{(Y-EY)^2}{(\sigma(Y))^2}\right)=$$

$$=\frac{E((X-EX)^2)}{DX}\pm2\frac{E((X-EX)(Y-EY))}{\sigma(X)\cdot\sigma(Y)}+\frac{E((Y-EY)^2)}{DY}=\frac{DX}{DX}\pm2\frac{\operatorname{cov}(X,Y)}{\sigma(X)\cdot\sigma(Y)}+\frac{DX}{DY}=$$

$=1\pm2\rho(X,Y)+1=2\pm2\rho(X,Y)$. В итоге доказаны два неравенства: $\begin{cases}2+2\rho(X,Y)\ge0,\\2-2\rho(X,Y)\ge0.\end{cases}$

Решив их получим $\begin{cases}\rho(X,Y)\ge-1,\\\rho(X,Y)\le1,\end{cases}$ что и доказывает требуемое утверждение $|\rho(X,Y)|\le1$.

2. Если $|\rho(X,Y)|=1$, то существуют такие числа a и b, что $Y=aX+b$, причём, если $\rho(X,Y)=1$, то $a>0$, а если $\rho(X,Y)=-1$, $a<0$.

   Разберём сначала случай $\rho(X,Y)=1$. Из доказательства предыдущего свойства следует, что $E\left(Z_1^2\right)=E\left(\left(\frac{X-EX}{\sigma(X)}-\frac{Y-EY}{\sigma(Y)}\right)^2\right)=2-2\rho(X,Y)=2-2\cdot1=0$. Из свойства дисперсии $DZ_1=E(Z_1^2)-(EZ_1)^2\Leftrightarrow DZ_1+(EZ_1)^2=E(Z_1^2)$ следует, что равна нулю сумма двух неотрицательных случайных величин: $DZ_1+(EZ_1)^2=0$. Поэтому обе эти величины равны нулю: $\begin{cases}DZ_1=0,\\(EZ_1)^2=0.\end{cases}\Leftrightarrow\begin{cases}Z_1=\mathrm{Const},\\EZ_1=0.\end{cases}\Leftrightarrow Z_1=0\Leftrightarrow$

$$\Leftrightarrow\frac{X-EX}{\sigma(X)}-\frac{Y-EY}{\sigma(Y)}=0\Leftrightarrow\frac{Y-EY}{\sigma(Y)}=\frac{X-EX}{\sigma(X)}\Leftrightarrow Y-EY=\frac{\sigma(Y)}{\sigma(X)}\cdot\left(X-EX\right)\Leftrightarrow$$

$$\Leftrightarrow Y=\frac{\sigma(Y)}{\sigma(X)}\cdot X-\frac{\sigma(Y)}{\sigma(X)}\cdot EX+EY\Leftrightarrow Y=aX+b,\ \text{где}\ a=\frac{\sigma(Y)}{\sigma(X)},\ b=EY-\frac{\sigma(Y)}{\sigma(X)}\cdot EX.$$

   Обратим внимание, что $a>0$ как отношение положительных чисел.

   Теперь рассмотрим ситуацию, когда $\rho(X,Y)=-1$. Аналогично следует:

$$DZ_2+(EZ_2)^2=E\left(Z_2^2\right)=E\left(\left(\frac{X-EX}{\sigma(X)}+\frac{Y-EY}{\sigma(Y)}\right)^2\right)=2+2\rho(X,Y)=2+2\cdot(-1)=0\Rightarrow$$

$$\Rightarrow \begin{cases} DZ_2 = 0, \\ (EZ_2)^2 = 0. \end{cases} \Leftrightarrow \begin{cases} Z_2 = \text{Const}, \\ EZ_2 = 0. \end{cases} \Leftrightarrow Z_2 = 0 \Leftrightarrow$$

$$\Leftrightarrow \frac{X - EX}{\sigma(X)} + \frac{Y - EY}{\sigma(Y)} = 0 \Leftrightarrow \frac{Y - EY}{\sigma(Y)} = -\frac{X - EX}{\sigma(X)} \Leftrightarrow Y - EY = -\frac{\sigma(Y)}{\sigma(X)} \cdot (X - EX) \Leftrightarrow$$

$\Leftrightarrow Y = -\frac{\sigma(Y)}{\sigma(X)} \cdot X + \frac{\sigma(Y)}{\sigma(X)} \cdot EX + EY \Leftrightarrow Y = aX + b$, теперь $a = -\frac{\sigma(Y)}{\sigma(X)}$, $b = EY + \frac{\sigma(Y)}{\sigma(X)} \cdot EX$. Заметим, что во втором случае a<0.

3. Если $Y=aX+b$, где $a \neq 0$, то $\rho(X,Y) = \begin{cases} 1, \text{ если } a > 0, \\ -1, \text{ если } a < 0. \end{cases}$ Для доказательства этого утверждения подставим данное выражение случайной величины Y через X в формулу, определяющую корреляцию: $\rho(X,Y) = \frac{\text{cov}(X,Y)}{\sigma(X)\cdot\sigma(Y)} =$

$$= \frac{E((X - EX)(Y - EY))}{\sqrt{DX} \cdot \sqrt{DY}} = \frac{E((X - EX)(aX + b - E(aX + b)))}{\sqrt{DX} \cdot \sqrt{D(aX + b)}} =$$

$$= \frac{E((X - EX)(aX + b - (aEX + b)))}{\sqrt{DX} \cdot \sqrt{D(aX) + Db}} = \frac{E((X - EX)(aX + b - aEX - b))}{\sqrt{DX} \cdot \sqrt{a^2 DX + 0}} =$$

$$= \frac{E((X - EX) \cdot a(X - EX))}{\sqrt{DX} \cdot \sqrt{a^2 DX}} = \frac{aE((X - EX)^2)}{\sqrt{DX} \cdot |a| \sqrt{DX}} = \frac{aDX}{|a|\,DX} = \frac{a}{|a|} = \begin{cases} \frac{a}{a}, \text{ если } a > 0, \\ \frac{a}{-a}, \text{ если } a < 0, \end{cases} =$$

$$= \begin{cases} 1, \text{ если } a > 0, \\ -1, \text{ если } a < 0. \end{cases}$$

Выведенные свойства позволяют сделать вывод о том, что именно показывает корреляция двух случайных величин. Не совсем верным будет утверждение, что это мера зависимости случайных величин. Корреляция отражает меру <u>линейности зависимости</u> случайных величин.

Неравенство Маркова

Пусть X неотрицательная случайная величина, у которой существует математическое ожидание EX. Пусть число t>0. Тогда $P(X \geq t) \leq \frac{EX}{t}$.

Доказательство. При доказательстве неотрицательности математического ожидания неотрицательной случайной величины было установлено, что при $X \geq 0$ верно $EX = \int_0^{+\infty} xf(x)dx$. Уменьшим промежуток интегрирования неотрицательной функции. По монотонности интеграла относительно промежутка получим: $EX = \int_0^{+\infty} xf(x)dx \geq \int_t^{+\infty} xf(x)dx$.

А теперь воспользуемся монотонностью интеграла по функции, уменьшив интегрируемую функцию: $EX = \int_0^{+\infty} xf(x)dx \geq \int_t^{+\infty} xf(x)dx \geq \int_t^{+\infty} tf(x)dx = t\int_t^{+\infty} f(x)dx = t \cdot P(X \geq t)$. Итак, $EX \geq t \cdot P(X \geq t) \Leftrightarrow P(X \geq t) \leq \frac{EX}{t}$.

Неравенство Чебышева

Пусть X случайная величина, у которой существует математическое ожидание EX и дисперсия DX. Тогда для любого числа $\varepsilon>0$ верно неравенство Чебышева: $P(|X-EX|\geq\varepsilon)\leq\dfrac{DX}{\varepsilon^2}$.

Доказательство. Рассмотрим неотрицательную случайную величину $Y=(X-EX)^2$. У неё существует математическое ожидание, более того $EY=DX$. По неравенству Маркова, если считать, что $t=\varepsilon^2$ (такое $t>0$), будем иметь: $P(Y\geq\varepsilon^2)\leq\dfrac{EY}{\varepsilon^2}\Leftrightarrow P\left((X-EX)^2\geq\varepsilon^2\right)\leq\dfrac{DX}{\varepsilon^2}\Leftrightarrow$

$$\Leftrightarrow P(|X-EX|\geq\varepsilon)\leq\frac{DX}{\varepsilon^2}=\frac{(\sigma(X))^2}{\varepsilon^2}.$$

Получается, что вероятность того, что отклонение случайной величины от её математического ожидания ограничена сверху на теоретическом уровне, независимо от вида распределения случайной величины. Заметим, что если $\varepsilon\leq\sigma(X)$, то $P(|X-EX|\geq\varepsilon)\leq\left(\dfrac{\sigma(X)}{\varepsilon}\right)^2\leq1$, что и так верно по ограниченности вероятности. То есть неравенство Чебышева даёт информацию только при $\varepsilon>\sigma(X)$. Если $\varepsilon=k\cdot\sigma(X)$, то $P(|X-EX|\geq k\cdot\sigma(X))\leq\left(\dfrac{\sigma(X)}{k\cdot\sigma(X)}\right)^2=\dfrac{1}{k^2}$. Ещё раз напомним, что это теоретическое ограничение на все случайные величины.

Для конкретных распределений (удобнее считать для непрерывных) ограничение будет жёстче:

$P(|X-EX|\geq k\cdot\sigma(X))=P(X-EX\geq k\cdot\sigma(X))+P(X-EX\leq-k\cdot\sigma(X))=P(X\geq EX+k\cdot\sigma(X))+P(X\leq EX-k\cdot\sigma(X))=$
$=1-P(X<EX+k\cdot\sigma(X))+P(X<EX-k\cdot\sigma(X))+P(X=EX-k\cdot\sigma(X))=1-F(EX+k\cdot\sigma(X))+F(EX-k\cdot\sigma(X))+0=$
$=1-F(EX+k\cdot\sigma(X))+F(EX-k\cdot\sigma(X))$.

Так для нормального распределения с параметрами $EX=a$, $DX=\sigma^2$ будем иметь:

$$P(|X-EX|\geq k\cdot\sigma(X))=1-F(a+k\cdot\sigma)+F(a-k\cdot\sigma)=1-\Phi\left(\frac{a+k\cdot\sigma-a}{\sigma}\right)+\Phi\left(\frac{a-k\cdot\sigma-a}{\sigma}\right)=$$

$$=1-\Phi\left(\frac{k\cdot\sigma}{\sigma}\right)+\Phi\left(\frac{-k\cdot\sigma}{\sigma}\right)=1-\Phi(k)+\Phi(-k)=1-\Phi(k)+1-\Phi(k)=2\left(1-\Phi(k)\right).$$

Для равномерного на отрезке $[a;b]$ распределения $EX=\dfrac{a+b}{2}$, $\sigma(X)=\dfrac{b-a}{2\sqrt{3}}$. Тогда $P(|X-EX|\geq k\cdot\sigma(X))=1-F\left(\dfrac{a+b}{2}+k\cdot\dfrac{b-a}{2\sqrt{3}}\right)+F\left(\dfrac{a+b}{2}-k\cdot\dfrac{b-a}{2\sqrt{3}}\right)$. Напомним, что функция распределения равномерного распределения задаётся несколькими формулами:

$$F(x)=\begin{cases}0,\text{если } x\leq a,\\ \dfrac{x-a}{b-a},\text{если } a<x<b,\\ 1,\text{если } x\geq b,\end{cases}$$

тогда

$$P(|X-EX|\geq k\cdot\sigma(X))=1-\begin{cases}0,\text{если}\dfrac{a+b}{2}+k\cdot\dfrac{b-a}{2\sqrt{3}}\leq a,\\ \dfrac{\dfrac{a+b}{2}+k\cdot\dfrac{b-a}{2\sqrt{3}}-a}{b-a},\text{если}\, a<\dfrac{a+b}{2}+k\cdot\dfrac{b-a}{2\sqrt{3}}<b,\\ 1,\text{если }\dfrac{a+b}{2}+k\cdot\dfrac{b-a}{2\sqrt{3}}\geq b,\end{cases}+$$

$$+\begin{cases}0,\text{если}\dfrac{a+b}{2}-k\cdot\dfrac{b-a}{2\sqrt{3}}\leq a,\\ \dfrac{\dfrac{a+b}{2}-k\cdot\dfrac{b-a}{2\sqrt{3}}-a}{b-a},\text{если}\, a<\dfrac{a+b}{2}-k\cdot\dfrac{b-a}{2\sqrt{3}}<b,\\ 1,\text{если }\dfrac{a+b}{2}-k\cdot\dfrac{b-a}{2\sqrt{3}}\geq b,\end{cases}=1-\begin{cases}0,\text{если }k\leq-\sqrt{3},\\ \dfrac{1}{2}+\dfrac{k}{2\sqrt{3}},\text{если}-\sqrt{3}<k<\sqrt{3},\\ 1,\text{если }k\geq\sqrt{3},\end{cases}+$$

$$+\begin{cases}0,\text{если }k\geq\sqrt{3},\\ \dfrac{1}{2}-\dfrac{k}{2\sqrt{3}},\text{если}-\sqrt{3}<k<\sqrt{3},\\ 1,\text{если }k\leq-\sqrt{3},\end{cases}$$ Учтём, что по смыслу задачи *k*>0, тогда

$$P(|X-EX|\geq k\cdot\sigma(X))=1-\begin{cases}\dfrac{1}{2}+\dfrac{k}{2\sqrt{3}},\text{если}\, k<\sqrt{3},\\ 1,\text{если }k\geq\sqrt{3},\end{cases}+\begin{cases}0,\text{если }k\geq\sqrt{3},\\ \dfrac{1}{2}-\dfrac{k}{2\sqrt{3}},\text{если}\, k<\sqrt{3},\end{cases}=$$

$$=1+\begin{cases}-\dfrac{1}{2}-\dfrac{k}{2\sqrt{3}}+\dfrac{1}{2}-\dfrac{k}{2\sqrt{3}},\text{если}\, k<\sqrt{3},\\ -1+0,\text{если }k\geq\sqrt{3},\end{cases}=1+\begin{cases}-\dfrac{k}{\sqrt{3}},\text{если}\, k<\sqrt{3},\\ -1,\text{если }k\geq\sqrt{3},\end{cases}=\begin{cases}1-\dfrac{k}{\sqrt{3}},\text{если}\, k<\sqrt{3},\\ 1-1,\text{если }k\geq\sqrt{3},\end{cases}=$$

$$=\begin{cases}1-\dfrac{k}{\sqrt{3}},\text{если }k<\sqrt{3},\\ 0,\text{если }k\geq\sqrt{3}.\end{cases}$$

Для экспоненциального распределения с параметром λ имеем $EX=1/\lambda$, $\sigma(X)=1/\lambda$. Тогда $P(|X-EX|\geq k\cdot\sigma(X))=1-F\left(\frac{1}{\lambda}+\frac{k}{\lambda}\right)+F\left(\frac{1}{\lambda}-\frac{k}{\lambda}\right)$. Функция распределения экспоненциального распределения тоже задаётся несколькими формулами: $F(x)=\begin{cases}0,\text{если}\, x\leq 0,\\ 1-e^{-\lambda x},\text{если}\, x>0,\end{cases}$ тогда

$$P(|X-EX|\geq k\cdot\sigma(X))=1-\begin{cases}0,\text{ если}\dfrac{1+k}{\lambda}\leq 0,\\ 1-e^{-(1+k)},\text{если }\dfrac{1+k}{\lambda}>0,\end{cases}+\begin{cases}0,\text{ если}\dfrac{1-k}{\lambda}\leq 0,\\ 1-e^{-(1-k)},\text{если }\dfrac{1-k}{\lambda}>0,\end{cases}=$$

$= 1 - \begin{cases} 0, \text{ если } k \le -1, \\ 1 - e^{-1-k}, \text{если } k > -1, \end{cases} + \begin{cases} 0, \text{ если } k \ge 1, \\ 1 - e^{k-1}, \text{если } k < 1. \end{cases}$ Учтём, что $k>0$, тогда

$$P(|X - EX| \ge k \cdot \sigma(X)) = 1 - (1 - e^{-1-k}) + \begin{cases} 0, \text{ если } k \ge 1, \\ 1 - e^{k-1}, \text{если } k < 1, \end{cases} = \begin{cases} e^{-1-k}, \text{ если } k \ge 1, \\ 1 + e^{-1-k} - e^{k-1}, \text{если } k < 1, \end{cases}$$

На рисунке 33 приведено сравнение теоретического ограничения (кривая 4) на величину $P(|X-EX| \ge k \cdot \sigma(X))$ с конкретными её зависимостями от числа k для нормального (кривая 2), равномерного (ломаная 3), экпоненциального (ломаная 1 из кривых) распределений.

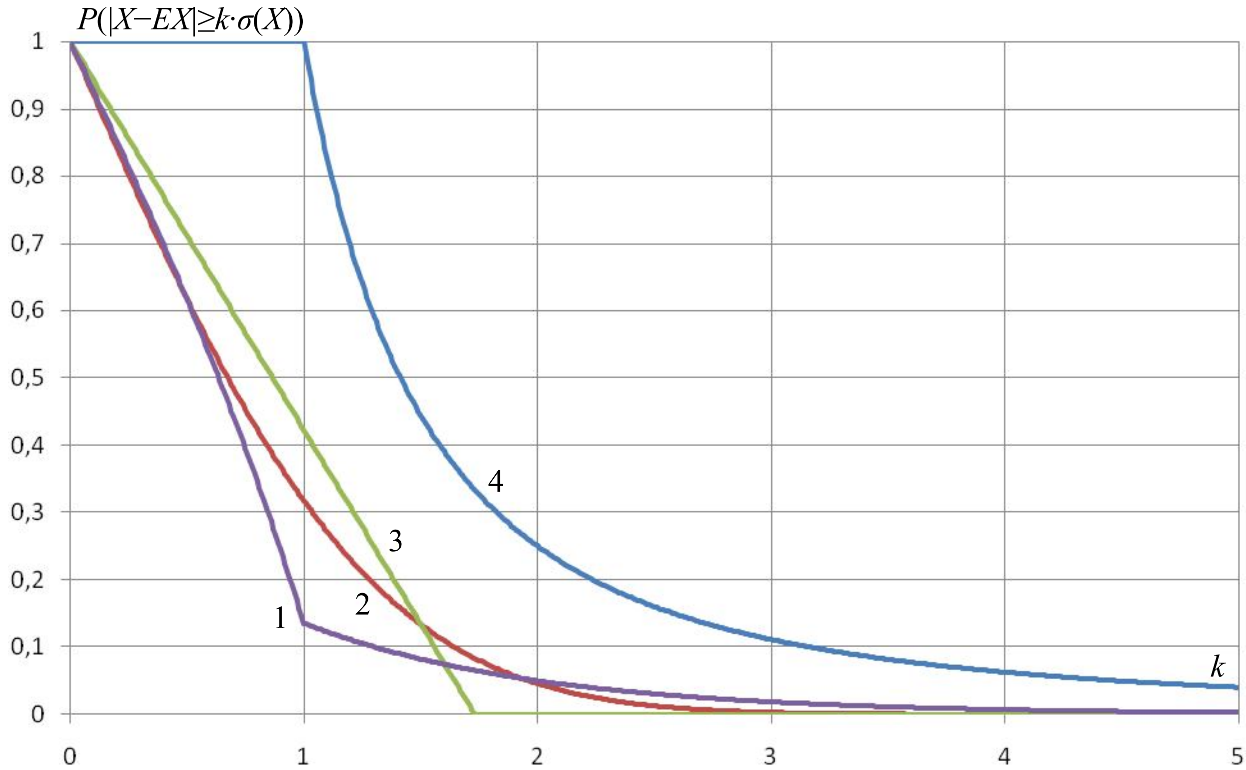

Рис. 33. Зависимости вероятности того, что случайная величина отклонится от своего математического ожидания в k раз больше её среднеквадратического отклонения в зависимости от k, для нормального (кривая 2), равномерного (ломаная 3), экпоненциального (ломаная 1 из кривых) распределений и теоретического ограничения, следующего из неравенства Чебышева (кривая 4).

Теорема Маркова (законы больших чисел)

Законами больших чисел называются теоремы, житейский смысл которых, заключается в том, что среднее арифметическое большого количества случайных величин практически не является случайным.

Теорема Маркова. Пусть X_1, X_2, … , X_n, … случайные величины (возможно даже зависимые) имеют математические ожидания $EX_1=a_1$, $EX_2=a_2$, … , $EX_n=a_n$, … . Пусть $\lim\limits_{n\to\infty}\left(\frac{1}{n^2}D\left(\sum\limits_{i=1}^{n}X_i\right)\right)=0$. Тогда для любого числа $\varepsilon>0$ верно: $\lim\limits_{n\to\infty}P\left(\left|\frac{1}{n}\sum\limits_{i=1}^{n}X_i-\frac{1}{n}\sum\limits_{i=1}^{n}a_i\right|\ge\varepsilon\right)=0$.

Доказательство. Назовём Y случайную величину, равную среднему арифметическому данных случайных величин: $Y=\frac{1}{n}\sum\limits_{i=1}^{n}X_i$. Заметим, что

$EY = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}E\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\left(EX_i\right) = \frac{1}{n}\sum_{i=1}^{n} a_i$. Также заметим, что $DY = D\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}D\left(\sum_{i=1}^{n} X_i\right)$. Применим для случайной величины Y неравенство Чебышева, попутно вспомнив, что вероятность всегда неотрицательное число: $0 \le P(|Y - EY| \ge \varepsilon) \le \frac{DY}{\varepsilon^2} = \frac{1}{\varepsilon^2}\cdot\frac{1}{n^2}D\left(\sum_{i=1}^{n} X_i\right)$. Перейдём в этом двойном неравенстве к пределу и по теореме о сжатой последовательности на основе того, что $\lim_{n\to\infty} 0 = 0$, как предел константы и $\lim_{n\to\infty}\left(\frac{1}{\varepsilon^2}\cdot\frac{1}{n^2}D\left(\sum_{i=1}^{n} X_i\right)\right) = \frac{1}{\varepsilon^2}\cdot\lim_{n\to\infty}\left(\frac{1}{n^2}D\left(\sum_{i=1}^{n} X_i\right)\right) = \frac{1}{\varepsilon^2}\cdot 0 = 0$, по условию, получим: $\lim_{n\to\infty} P\left(\left|Y - \frac{1}{n}\sum_{i=1}^{n} a_i\right| \ge \varepsilon\right) = 0$. Теорема доказана.

Заметим, что при применении теоремы Маркова не следует пренебрегать условием $\lim_{n\to\infty}\left(\frac{1}{n^2}D\left(\sum_{i=1}^{n} X_i\right)\right) = 0$ и использовать её для каких угодно зависимых случайных величин (вообще говоря, никогда не следует пренебрегать условиями теорем при их применении). Так, например, если все $X_i = X_1 \ne \text{Const}$, то $\lim_{n\to\infty}\left(\frac{1}{n^2}D\left(\sum_{i=1}^{n} X_i\right)\right) = \lim_{n\to\infty}\left(\frac{1}{n^2}D\left(\sum_{i=1}^{n} X_1\right)\right) = \lim_{n\to\infty}\left(\frac{1}{n^2}D\left(nX_1\right)\right) = \lim_{n\to\infty}\left(\frac{1}{n^2}n^2 DX_1\right) = \lim_{n\to\infty}\left(DX_1\right) = DX_1 \ne 0.$ Этот предел не равен нулю. И теорема Маркова неприменима.

Заметим, что утверждаемый предел в теореме Маркова можно сформулировать и иначе: $\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} a_i\right| < \varepsilon\right) = 1$.

Следствия из теоремы Маркова

На практике очень часто складываются независимые случайные величины. Используя эту независимость, можно упростить условие на дисперсию, сделав его легкопроверяемым. Сформулируем теоремы, которые формально следуют из теоремы Маркова.

Обобщённая теорема Чебышева. Пусть X_1, X_2, … , X_n, … попарно независимые (для взаимно независимых теорема тем более справедлива) случайные величины имеют математические ожидания $EX_1=a_1$, $EX_2=a_2$, … , $EX_n=a_n$, … . Пусть существует число $L>0$ (единое для всех случайных величин) ограничивающее дисперсии всех данных случайных величин: $DX_i<L$. Тогда для любого числа $\varepsilon>0$ верно: $\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} a_i\right| \ge \varepsilon\right) = 0$.

Доказательство. Докажем недостающее условие теоремы Маркова, благодаря независимости случайных величин дисперсия суммы равна сумме дисперсий: $0 \le \lim_{n\to\infty}\left(\frac{1}{n^2}D\left(\sum_{i=1}^{n} X_i\right)\right) = \lim_{n\to\infty}\left(\frac{1}{n^2}\sum_{i=1}^{n} DX_i\right) \le \lim_{n\to\infty}\left(\frac{1}{n^2}\sum_{i=1}^{n} L\right) = \lim_{n\to\infty}\left(\frac{1}{n^2}nL\right) = \lim_{n\to\infty}\frac{L}{n} = 0$.

Следовательно, $\lim_{n\to\infty}\left(\frac{1}{n^2}D\left(\sum_{i=1}^{n} X_i\right)\right) = 0$ и обобщённая теорема Чебышева справедлива, как следствие теоремы Маркова.

Теорема Чебышева. Пусть X_1, X_2, … , X_n, … попарно независимые (для взаимно независимых теорема тем более справедлива) одинаково распределённые случайные величины, имеющие математические ожидания $EX_1=EX_2=\ldots=EX_n=\ldots=a$ и дисперсии. Тогда для любого числа $\varepsilon>0$ верно: $\lim\limits_{n\to\infty} P\left(\left|\frac{1}{n}\sum\limits_{i=1}^{n} X_i - a\right| \ge \varepsilon\right) = 0$.

Доказательство. Поскольку DX_i число, то любое число, превосходящее его может быть выбрано в качестве числа L в обобщённой теорема Чебышева. Также $a=\frac{1}{n}\sum\limits_{i=1}^{n} a_i$. И теорема Чебышева справедлива, как следствие обобщённой теоремы Чебышева.

Сформулируем ещё пару теорем, формально не являющихся законами больших чисел, поскольку в них идёт речь не о случайных величинах, а о случайных событиях, но тем не менее, непосредственно вытекающих из обобщённой теоремы Чебышева.

Теорема Пуассона. Пусть проводится последовательность независимых (можно попарно независимых) испытаний с двумя исходами (успехом и неудачей) с вероятностью успеха в i-ом испытании равном p_i. Тогда для любого числа $\varepsilon>0$ верно: $\lim\limits_{n\to\infty} P\left(\left|f_n - \frac{1}{n}\sum\limits_{i=1}^{n} p_i\right| \ge \varepsilon\right) = 0$, где f_n частота появления успеха в первых n испытаниях.

Для доказательства введём случайную величину X_i, равную числу успехов в отдельно взятом одном i-ом испытании. Эта случайная величина имеет распределение Бернулли с параметром p_i. Её математическое ожидание $EX_i=p_i$. Дисперсия $DX_i=p_i(1-p_i)\le 1/4<1$. А частота $f_n=\frac{1}{n}\sum\limits_{i=1}^{n} X_i$. Тогда доказываемое утверждение действительно является частным случаем обобщённой теоремы Чебышева.

Теорема Бернулли. Пусть проводится последовательность независимых (можно попарно независимых) испытаний с двумя исходами (успехом и неудачей) с вероятностью успеха p в каждом испытании. Тогда для любого числа $\varepsilon>0$ верно: $\lim\limits_{n\to\infty} P\left(\left|f_n - p\right| \ge \varepsilon\right) = 0$, где f_n частота появления успеха в первых n испытаниях.

Очевидно, что эта теорема является следствием теоремы Пуассона при $p_i=p$, поскольку $\frac{1}{n}\sum\limits_{i=1}^{n} p = \frac{1}{n} np = p$.

Смысл теоремы Бернулли уже выяснялся во время обсуждения того, что такое вероятность. А теорема Пуассона объясняет свойство устойчивости частот не только для идеально проводимых экспериментов, а и тогда, когда вероятность события изменяется от опыта к опыту, лишь бы была независимость испытаний.

Случайные блуждания

Рассмотрим самую простейшую бесконечную серию независимых испытаний с двумя исходами (успехом и неудачей) — частный случай, когда вероятность успеха равна вероятности неудачи и равна ½ (бросания симметричной монеты). Назовём выигрышем Z разность между числом успехов и числом неудач. Если число испытаний равно n, число успехов m, то число неудач $(n-m)$, и выигрыш составит $Z=m-(n-m)=2m-n$. К сожалению, обыватели почему-то думают, что Z «приближается» к нулю. Попытаемся разубедить в этом читателя.

Сначала посмотрим, что следует из теоремы Бернулли. Для любого числа $\delta>0$ верно:

$$\lim_{n\to\infty} P\left(\left|\frac{m}{n}-\frac{1}{2}\right| \ge \varepsilon\right) = 0 \Leftrightarrow \lim_{n\to\infty} P\left(\left|\frac{2m-n}{2n}\right| \ge \varepsilon\right) = 0 \Leftrightarrow \lim_{n\to\infty} P\left(\left|\frac{Z}{2n}\right| \ge \varepsilon\right) = 0 \Leftrightarrow \lim_{n\to\infty} P\left(\frac{|Z|}{n} \ge 2\varepsilon\right) = 0.$$

Итак, к нулю «приближается» (стремится по вероятности) не Z, а дробь $\frac{Z}{n}$. Знаменатель этой дроби стремится к бесконечности, и чтобы дробь стремилась к нулю, вовсе нет никакой необходимости в стремлении к нулю числителя. Числитель может даже стремиться к бесконечности! Например, как $\sqrt{n}$. Итак, никаких теоретических предпосылок в стремлении выигрыша Z к нулю нет.

Решим следующий вопрос. Какова вероятность того, что хотя бы в один момент времени выигрыш будет равен одному? Обозначим p_n вероятность того, что хотя бы в один момент времени выигрыш составит n. Заметим, что в таких обозначениях искомой величиной является p_1. Заметим также, что p_0=1. На рисунке 34 изображена схема начала эксперимента. На ней точки ветвления обозначены вероятностями получить выигрыш, равный одному, хотя бы в один момент времени, при движении по схеме из данной точки. Если была неудача, то для достижения выигрыша равного одному, успехов требуется уже на один больше.

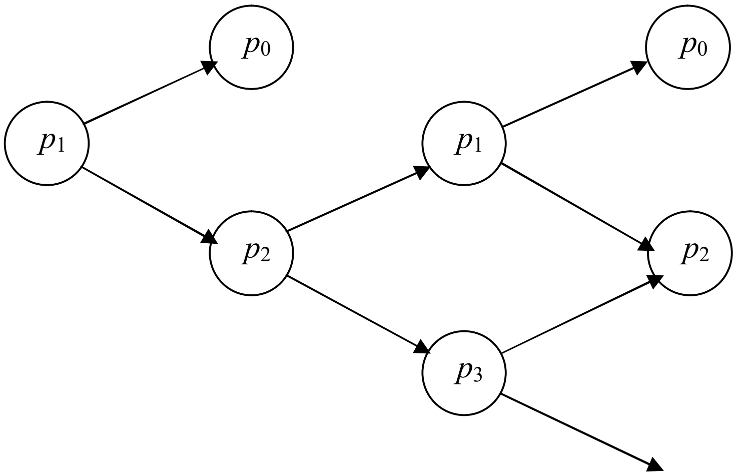

Рис. 34. Схема эксперимента по бесконечному подбрасыванию монеты до выигрыша, равного одному.

Для первого шага по формуле полной вероятности, учитывая, что p_0=1, получим: $p_1 = \frac{1}{2}\cdot 1 + \frac{1}{2}\cdot p_2 \Leftrightarrow p_2 = 2p_1 - 1$. Для второго шага: $p_2 = \frac{1}{2}\cdot p_1 + \frac{1}{2}\cdot p_3 \Leftrightarrow p_3 = 2p_2 - p_1$. Если подставить выражение для p_2 во вторую формулу, получим: $p_3 = 2(2p_1 - 1) - p_1 = 3p_1 - 2$. Возможно, Вы уже обнаружили закономерность: $p_n = np_1 - n + 1$. Докажем эту формулу методом математической индукции. Предположим, что доказываемое утверждение верно для двух, идущих подряд натуральных чисел: ($n-1$) и n. И докажем, что в этом случае подобное утверждение верно и для числа (n+1). Итак, дано $p_{n-1} = (n-1)p_1 - (n-1) + 1$ и $p_n = np_1 - n + 1$, а надо доказать, что $p_{n+1} = (n+1)p_1 - (n+1) + 1$. Из формулы полной вероятности следует $p_n = \frac{1}{2}\cdot p_{n-1} + \frac{1}{2}\cdot p_{n+1} \Leftrightarrow p_{n+1} = 2p_n - p_{n-1} =$ $= 2(np_1 - n + 1) - ((n-1)p_1 - (n-1) + 1) = np_1 + p_1 - n = (n+1)p_1 - (n+1) + 1$. Выше было непосредственно получено, что доказываемое утверждение справедливо для n=1 и n=2, следовательно оно справедливо и для n=3. Тогда оно будет справедливо для n=2 и n=3, и поэтому будет верно при n=4, и так далее. Теперь выразим искомую вероятность p_1 из доказанной формулы: $p_n = np_1 - n + 1 \Leftrightarrow p_1 = \frac{p_n + n - 1}{n} = 1 - \frac{1 - p_n}{n}$. Перейдём в полученном равенстве к пределу: $\lim_{n\to\infty} p_1 = \lim_{n\to\infty}\left(1 - \frac{1 - p_n}{n}\right)$. Учитывая, что предел константы равен самой константе, а предел произведения ограниченной ($p_n \in [0;1]$, как вероятность) последовательности на бесконечно малую равен нулю, получим: p_1=1−0=1. Оказывается рано или поздно выигрыш обязательно будет равен одному. Обратим внимание на то, что $p_n = np_1 - n + 1 = n\cdot 1 - n + 1 = 1$. То есть, рано или поздно выигрыш (а в силу симметрии задачи и проигрыш) составит любое число (в том числе и любое сколь угодно большое)! Вряд ли после этого у кого-то осталось впечатление, что с ростом числа испытаний выигрыш приближается к нулю. Как будет вести себя зависимость выигрыша Z от числа испытаний n? Учитывая, что все испытания независимы (монета не имеет памяти) можно утверждать, что для монеты каждый раз всё как будто начинается с нулевого выигрыша. И зависимость Z от n как будто каждый раз начинается сначала. Конечно, неблагодарное занятие угадывать, как упадёт монета, но можно спрогнозировать, что если сгладить отдельные шероховатости графика, то зависимость выигрыша Z от числа испытаний n будет похожа на осциллирующую функцию, у которой увеличивается и амплитуда и период колебаний.

Введём в рассмотрение случайную величину X_i, равную выигрышу в i-ом испытании. Это дискретная случайная величина, таблица распределения которой:

$X_{i,j}$:	-1	1
p_j:	½	½

Математическое ожидание $EX_i=-1\cdot½+1\cdot½=0$. Найдём её дисперсию.

$X_{i,j}$:	-1	1
p_j:	½	½
$X_{i,j}-EX_i$	-1	1
$(X_{i,j}-EX_i)^2$	1	1

$DX_i=1\cdot½+1\cdot½=1$. Заметим, что $Z=\sum_{i=1}^{n}X_i$. Найдём $EZ=E\left(\sum_{i=1}^{n}X_i\right)=\sum_{i=1}^{n}EX_i=\sum_{i=1}^{n}0=n\cdot 0=0$ и $DZ=D\left(\sum_{i=1}^{n}X_i\right)=\sum_{i=1}^{n}DX_i=\sum_{i=1}^{n}1=n\cdot 1=n$. И среднеквадратическое отклонение $\sigma(Z)=\sqrt{DZ}=\sqrt{n}$. Физически невозможен выход Z за промежуток $[-n;n]$. Согласно правилу трёх сигм практически невозможен выход случайной величины Z из промежутка $[-3\sqrt{n};3\sqrt{n}]$. А неравенство Чебышева не накладывает никаких ограничений на нахождение случайной величины Z в промежутке $[-\sqrt{n};\sqrt{n}]$. На рисунке 35 показан график зависимости выигрыша от числа испытаний и линии, на которых отклонение от математического ожидания в целое число раз (в один, два или три) превосходит среднеквадратическое.

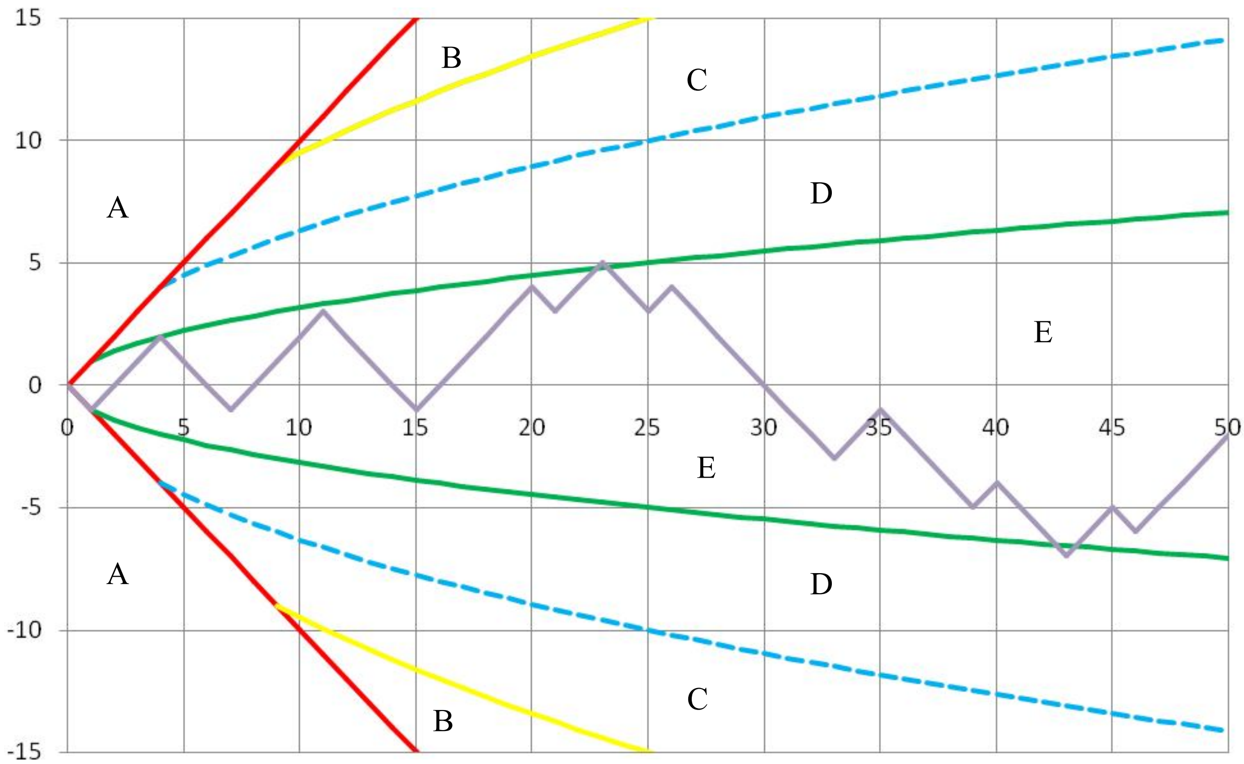

Рис.35. Зависимость выигрыша от числа испытаний при равенстве вероятностей успеха и неудачи (получена в компьютерном эксперименте). Линии разделяют зоны, где точке графика находиться: A — невозможно; B — практически невозможно (отклонение превзошло три среднеквадратических); C — крайне удивительно (отклонение превзошло два среднеквадратических); D — несколько удивительно (отклонение превзошло среднеквадратическое); E — совершенно нормально.

Центральная предельная теорема

Центральной предельной теоремой называют утверждение о том, что при выполнении некоторых условий распределение суммы случайных величин стремится к нормальному закону при стремлении числа слагаемых к бесконечности. Сформулируем без доказательств два варианта центральной предельной теоремы.

Теорема Ляпунова. Пусть X_1, X_2, … , X_n, … взаимно независимые случайные величины, имеющие математические ожидания EX_1, EX_2, … , EX_n, … и дисперсии DX_1, DX_2, … , DX_n, … . Пусть случайная величина $Y_n = \sum_{i=1}^{n} X_i$, равна сумме первых n случайных величин. Найдём $a_n = EY_n = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} EX_i$ и $\sigma_n = \sigma(Y_n) = \sqrt{DY_n} = \sqrt{D\left(\sum_{i=1}^{n} X_i\right)} = \sqrt{\sum_{i=1}^{n} DX_i}$ её математическое ожидание и среднеквадратическое отклонение. Пусть ещё $\lim\limits_{n\to\infty} \dfrac{\sum_{i=1}^{n} E\left(|X_i - EX_i|^3\right)}{(\sigma_n)^3} = 0$. Тогда случайная величина Z_n, полученная из Y_n нормированием и центрированием, то есть: $Z_n = \dfrac{Y_n - a_n}{\sigma_n}$ стремится к нормальной, у которой математическое ожидание равно нулю, а дисперсия единице. Если функцию распределения случайной величины Z_n обозначить $F_n(x)$, то для любого x выполняется: $\lim\limits_{n\to\infty} F_n(x) = \Phi(x)$.

Теорема Леви. Пусть X_1, X_2, … , X_n, … взаимно независимые одинаково распределённые случайные величины имеющие математические ожидания, обозначенные $a=EX_1=EX_2=\ldots=EX_n=\ldots$, и дисперсии, обозначенные $\sigma^2=DX_1=DX_2=\ldots=DX_n=\ldots$. Пусть случайная величина $Y_n = \sum_{i=1}^{n} X_i$, равна сумме первых n случайных величин. Найдём $a_n = EY_n = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} EX_i = \sum_{i=1}^{n} a = na$ и $\sigma_n = \sigma(Y_n) = \sqrt{DY_n} = \sqrt{D\left(\sum_{i=1}^{n} X_i\right)} = \sqrt{\sum_{i=1}^{n} DX_i} = \sqrt{\sum_{i=1}^{n} \sigma^2} = \sqrt{n\sigma^2} = \sigma\sqrt{n}$ её математическое ожидание и среднеквадратическое отклонение. Тогда случайная величина Z_n, полученная из Y_n нормированием и центрированием, то есть: $Z_n = \dfrac{Y_n - na}{\sigma\sqrt{n}}$ стремится к нормальной, у которой математическое ожидание равно нулю, а дисперсия единице. Если функцию распределения случайной величины Z_n обозначить $F_n(x)$, то для любого x выполняется: $\lim\limits_{n\to\infty} F_n(x) = \Phi(x)$.

Теперь ясно, что появление функции $\Phi(x)$ в интегральной теореме Муавра−Лапласа имеет причины, непосредственно никак не связанные с фактически рассматриваемым в этой теореме распределением Бернулли.

На практике невозможно просуммировать бесконечное количество случайных величин, но если ни одна из них не вносит превалирующего вклада, то распределение суммы приближённо можно считать нормальным. Именно поэтому большинство случайных величин в биологии распределены по законам, близким к нормальному.

Многомерные случайные величины (случайные векторы)

Если n упорядоченных случайных величин (неважно зависимых или нет) определены на одном вероятностном пространстве, то можно считать, что это не n величин, а одна, но n-мерная. Формально из них можно образовать вектор размерности n. И если считать, что случайная величина это функция, заданная на пространстве элементарных событий, то *случайный вектор*, это многомерная функция от одного аргумента: случайного события.

В простейшем случае количество компонент случайного вектора равно двум, то есть рассматривается упорядоченная пара (X,Y) случайных величин, то есть двумерная случайная величина.

Назовём её функцией распределения (это будет функция двух аргументов) вероятность $F(x,y)=P((X<x)\cap(Y<y))$. Такая вероятность уже изучалась во время выяснения вопросов, связанных с независимостью двух случайных величин. Тогда же было показано, что вероятность попадания двумерной точки (X,Y) в прямоугольник $[a;b)\times[c;g)$, стороны которого параллельны осям координат, даже при зависимых случайных величинах, равна $P((a\le X<b)\cap(c\le Y<g))=P((X<b)\cap(Y<g))-P((X<b)\cap(Y<c))-P((X<a)\cap(Y<g))+P((X<a)\cap(Y<c))$. С помощью понятия функции распределения это равенство можно записать так: $P((a\le X<b)\cap(c\le Y<g))=F(b,g)-F(b,c)-F(a,g)+F(a,c)$.

Для функции распределения двумерной случайной величины выполняются свойства, аналогичные свойствам обыкновенной функции распределения:

1. $0\le F(x,y)\le 1$.
2. $\lim\limits_{\substack{x\to-\infty\\ y\to-\infty}} F(x,y)=P((X<-\infty)\cap(Y<-\infty))=P(\varnothing\cap\varnothing)=P(\varnothing)=0$.

3. $\lim\limits_{\substack{x\to+\infty\\y\to-\infty}} F(x,y)=P((X<+\infty)\cap(Y<-\infty))=P(\Omega\cap\varnothing)=P(\varnothing)=0.$
4. $\lim\limits_{y\to-\infty} F(x,y)=P((X<x)\cap(Y<-\infty))=P((X<x)\cap\varnothing)=P(\varnothing)=0.$
5. $\lim\limits_{\substack{x\to-\infty\\y\to+\infty}} F(x,y)=P((X<-\infty)\cap(Y<+\infty))=P(\varnothing\cap\Omega)=P(\varnothing)=0.$
6. $\lim\limits_{x\to-\infty} F(x,y)=P((X<-\infty)\cap(Y<y))=P(\varnothing\cap(Y<y))=P(\varnothing)=0.$
7. $\lim\limits_{\substack{x\to+\infty\\y\to+\infty}} F(x,y)=P((X<+\infty)\cap(Y<+\infty))=P(\Omega\cap\Omega)=P(\Omega)=1.$
8. $\lim\limits_{y\to+\infty} F(x,y)=P((X<x)\cap(Y<+\infty))=P((X<x)\cap\Omega)=P(X<x)=F_1(x)$, где $F_1(x)$ функция распределения одномерной случайной величины X.
9. $\lim\limits_{x\to+\infty} F(x,y)=P((X<+\infty)\cap(Y<y))=P(\Omega\cap(Y<y))=P(Y<y)=F_2(y)$, где $F_2(y)$ функция распределения одномерной случайной величины Y.
10. Функция распределения $F(x,y)$ неубывающая по каждому из аргументов, то есть из $x_1<x_2$ следует, что $F(x_1,y)\le F(x_2,y)$, а из $y_1<y_2$ следует, что $F(x,y_1)\le F(x,y_2)$.

Плотность вероятности двумерной случайной величины

Назовём плотностью вероятности двумерной случайной величины вторую частную смешанную производную функции распределения: $f(x,y)=\dfrac{\partial^2 F(x,y)}{\partial y\,\partial x}=F''_{xy}(x,y)$.

Аналогично одномерному случаю $f(x,y)\ge 0$. Заметим, что вероятность попадания двумерной случайной величины в прямоугольник $D=[a;b)\times[c;g)$, стороны которого параллельны осям координат, может быть записана в виде двойного интеграла от плотности вероятности, по этой прямоугольной области:

$$\iint\limits_{[a;b)\times[c;g)} f(x,y)dD=\int_a^b dx\int_c^g f(x,y)dy=\int_a^b dx\int_c^g \left(F_x'(x,y)\right)_y' dy=\int_a^b dx\, F_x'(x,y)\Big|_c^g=$$

$$=\int_a^b \left(F_x'(x,g)-F_x'(x,c)\right)dx=\int_a^b F_x'(x,g)dx-\int_a^b F_x'(x,c)dx=F(x,g)\Big|_a^b-F(x,c)\Big|_a^b=$$

$$=F(b,g)-F(a,g)-\left(F(b,c)-F(a,c)\right)=F(b,g)-F(a,g)-F(b,c)+F(a,c)=$$

$=P((a\le X<b)\cap(c\le Y<g))$. Поскольку прямоугольник D произвольный (в частности сколь угодно маленький), то из определения двойного интеграла следует, что вероятность попадания двумерной случайной величины в область D произвольной формы (которую можно представить в виде объединения бесконечного количества непересекающихся прямоугольников) тоже будет равна двойному интегралу от плотности вероятности по области D: $P((X,Y)\in D)=\iint\limits_D f(x,y)dD$. Прямоугольник $D=[a;b)\times[c;g)$ может быть и бесконечных размеров, что приводит к равенствам:

$$\int_{-\infty}^{+\infty} dx\int_{-\infty}^{+\infty} f(x,y)dy=\iint\limits_{(-\infty;+\infty)\times(-\infty;+\infty)} f(x,y)dD=P((-\infty\le X<+\infty)\cap(-\infty\le Y<+\infty))=P(\Omega\cap\Omega)=P(\Omega)=1,$$

$$\int_{-\infty}^{x} dt\int_{-\infty}^{+\infty} f(t,y)dy=\iint\limits_{(-\infty;x)\times(-\infty;+\infty)} f(t,y)dD=P((-\infty\le X<x)\cap(-\infty\le Y<+\infty))=P((X<x)\cap\Omega)=$$

$$=P(X<x)=F_1(x),$$

$$\int_{-\infty}^{+\infty} dx \int_{-\infty}^{y} f(x,t)dt = \iint_{(-\infty;+\infty)\times(-\infty;y)} f(x,t)dD = P((-\infty \le X < +\infty)\cap(-\infty \le Y < y)) = P(\Omega \cap (Y < y)) =$$

$$= P(Y<y) = F_2(y), \quad \int_{-\infty}^{x} dt \int_{-\infty}^{y} f(t,v)dv = \iint_{(-\infty;x)\times(-\infty;y)} f(t,v)dD = P((-\infty \le X < x)\cap(-\infty \le Y < y)) = F(x,y)$$

.

Найдём плотности вероятности каждого из компонент вектора (X,Y). Плотность вероятности одномерной случайной величины X обозначим $f_1(x)$. Согласно определению она равна: $f_1(x) = F_1'(x) = \left(F_1(x)\right)'_x = \left(\int_{-\infty}^{x} dt \int_{-\infty}^{+\infty} f(t,y)dy\right)'_x = \int_{-\infty}^{+\infty} f(x,y)dy$. Последнее равенство записано на основании теоремы Барроу (о производной от интеграла по его верхнему пределу). Аналогично можно обозначить $f_2(y)$ плотность вероятности одномерной случайной величины Y. Тогда $f_2(y) = F_2'(y) = \left(F_2(y)\right)'_y = \left(\int_{-\infty}^{+\infty} dx \int_{-\infty}^{y} f(x,t)dt\right)'_y = \int_{-\infty}^{+\infty} f(x,y)dx$.

Независимость компонент случайного вектора

Если компоненты случайного вектора независимые, то можно получить правило произведения для функций распределения: $F(x,y)=P((X<x)\cap(Y<y))=P(X<x)\cdot P(Y<y)=$ $=F_1(x)\cdot F_2(y)$. Интересно, что при независимости компонент случайного вектора, правило произведения будет справедливо и для плотностей вероятности: $f(x,y) = \left(F(x,y)\right)''_{xy} = \left(\left(F_1(x)\cdot F_2(y)\right)'_x\right)'_y = \left(F_2(y)\cdot\left(F_1(x)\right)'_x\right)'_y = \left(F_2(y)\cdot f_1(x)\right)'_y = f_1(x)\cdot\left(F_2(y)\right)'_y =$ $= f_1(x)\cdot f_2(y)$.

Двумерное нормальное распределение

Двумерное распределение будем называть нормальным, если её плотность вероятности выражается формулой: $f(x,y) = \dfrac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\cdot\exp\left(\dfrac{1}{(1-\rho^2)}\cdot\left(-\dfrac{(x-a_1)^2}{2\sigma_1^2}+\dfrac{\rho(x-a_1)(y-a_2)}{\sigma_1\sigma_2}-\dfrac{(y-a_2)^2}{2\sigma_2^2}\right)\right)$. У этого закона есть <u>пять</u> параметров: a_1, a_2, σ_1, σ_2, ρ. Их смысл можно выяснить, вычислив плотности вероятностей отдельных компонент. Примем без доказательства, что $f_1(x) = \int_{-\infty}^{+\infty} f(x,y)dy = \ldots = \dfrac{1}{\sigma_1\sqrt{2\pi}}\cdot\exp\left(-\dfrac{(x-a_1)^2}{2\sigma_1^2}\right) = \dfrac{1}{\sigma_1}\cdot\varphi\left(\dfrac{x-a_1}{\sigma_1}\right)$. Аналогично: $f_2(y) = \int_{-\infty}^{+\infty} f(x,y)dx = \ldots = \dfrac{1}{\sigma_2\sqrt{2\pi}}\cdot\exp\left(-\dfrac{(y-a_2)^2}{2\sigma_2^2}\right) = \dfrac{1}{\sigma_2}\cdot\varphi\left(\dfrac{y-a_2}{\sigma_2}\right)$. Следовательно, каждая из компонент этого двумерного вектора распределена по обыкновенному одномерному нормальному закону с соответствующими параметрами: $EX=a_1$, $\sigma(X)=\sigma_1$, $EY=a_2$, $\sigma(Y)=\sigma_2$. Для выяснения смысла параметра ρ, необходимо найти ковариацию случайных величин X и Y. Опуская вычисление двойного несобственного интеграла, будем иметь: $\mathrm{cov}(X,Y) = \int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} (x-a_1)(y-a_2)f(x,y)dy = \ldots = \rho\sigma_1\sigma_2$. Поэтому параметр ρ совпадает с корреляцией компонент случайного вектора: $\rho = \dfrac{\mathrm{cov}(X,Y)}{\sigma_1\cdot\sigma_2} = \dfrac{\mathrm{cov}(X,Y)}{\sigma(X)\cdot\sigma(Y)} = \rho(X,Y)$.

Если случайные величины X и Y независимые, то корреляция $\rho=0$ и очевидно выполняется правило произведения для плотностей вероятностей: $f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2}\cdot\exp\left(-\frac{(x-a_1)^2}{2\sigma_1^2}-\frac{(y-a_2)^2}{2\sigma_2^2}\right)=\frac{1}{\sigma_1\sigma_2}\cdot\varphi\left(\frac{x-a_1}{\sigma_1}\right)\cdot\varphi\left(\frac{y-a_2}{\sigma_2}\right)=f_1(x)\cdot f_2(y)$.

При выяснении свойств корреляции было установлено, что если корреляция равна нулю, то это ещё не означает, что случайные величины независимые. Однако, из текущего пункта следует, что при наличии информации о том, что случайные величины распределены нормально (а на практике очень часто приходится иметь дело со случайными величинами, распределения которых близки к нормальному закону), равенство корреляции нулю однозначно доказывает независимых случайных величин.

Правило свёртки

Рассмотрим задачу о нахождении вида распределения суммы двух заданных случайных величин X и Y. Обозначим $Z=X+Y$. Составим из заданных случайных величин двумерную случайную величину: (X,Y). Пусть плотность распределения этой двумерной случайной величины равна $f(x,y)$. Если исходные одномерные случайные величины являются независимыми (как это часто бывает на практике) и имеют плотности вероятности $f_1(x)$ и $f_2(y)$, то плотность вероятности двумерной величины будет равна $f(x,y)=f_1(x)\cdot f_2(y)$. В случае зависимости исходных случайных величин X и Y о плотности вероятности $f(x,y)$ ничего определённого сказать нельзя.

Найдём функцию распределения случайной величины Z. По определению функции распределения она равна $F(x)=P(Z<x)=P(X+Y<x)$. Такая вероятность может быть вычислена как двойной интеграл $P(X+Y<x)=\iint\limits_D f(u,v)dD$, где D это область на координатной плоскости OXY, соответствующая неравенству $X+Y<x$, причём x рассматривается как константа. Такому неравенству соответствует полуплоскость (бесконечная), находящаяся левее и ниже прямой $X+Y=x\Leftrightarrow Y=x-X$ (см. рис. 36).

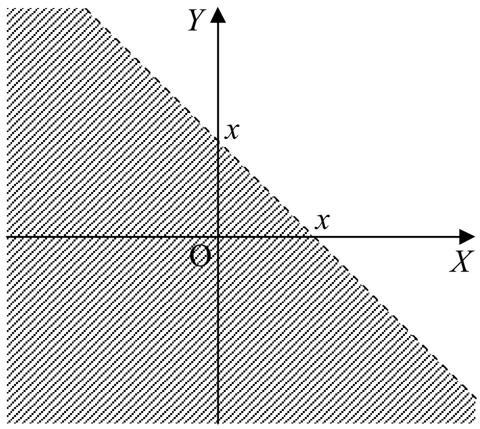

Рис. 36. Область, соответствующая неравенству $X+Y<x$.

Для нахождения двойного интеграла расставим пределы интегрирования, переходя к повторному интегралу: $F(x)=P(X+Y<x)=\int\limits_{-\infty}^{+\infty}du\int\limits_{-\infty}^{x-u}f(u,v)dv$.

Найдём плотность вероятности случайной величины Z, продифференцировав функцию распределения с помощью теоремы Барроу (о производной интеграла по верхнему пределу): $f_Z(x)=F'(x)=\left(\int\limits_{-\infty}^{+\infty}dt\int\limits_{-\infty}^{x-u}f(u,v)dv\right)'_x=$

$=\int\limits_{-\infty}^{+\infty}\left(\int\limits_{-\infty}^{x-u}f(u,v)dv\right)'_x du=\int\limits_{-\infty}^{+\infty}f(u,x-u)du$. Задача решена, поскольку найденная плотность вероятности однозначно определяет случайную величину.

Заметим, что если бы при переходе от двойного интеграла к повторному интегралу, пределы интегрирования были бы расставлены в другом порядке, то получилась бы другая (симметричная первой) формула:

$f_Z(x)=F'(x)=\left(\int\limits_{-\infty}^{+\infty}dv\int\limits_{-\infty}^{x-y}f(u,v)du\right)'_x=\int\limits_{-\infty}^{+\infty}\left(\int\limits_{-\infty}^{x-v}f(u,v)du\right)'_x dv=\int\limits_{-\infty}^{+\infty}f(x-v,v)dv$. Попутно доказано тождество: $\int\limits_{-\infty}^{+\infty}f(t,x-t)dt=\int\limits_{-\infty}^{+\infty}f(x-t,t)dt$.

Если заданные случайные величины X и Y независимые, то, как уже упоминалось $f(x,y)=f_1(x)\cdot f_2(y)$, и тогда $f_Z(x)=\int\limits_{-\infty}^{+\infty}f_1(x-t)f_2(t)dt$ или $f_Z(x)=\int\limits_{-\infty}^{+\infty}f_1(t)f_2(x-t)dt$. В математике существует операция над функциями, называемая *свёртка*. Она обозначается звёздочкой и $f_1(x)*f_2(x)=\int\limits_{-\infty}^{+\infty}f_1(x-t)f_2(t)dt=\int\limits_{-\infty}^{+\infty}f_1(t)f_2(x-t)dt$. Причём корректность этого определения (в том смысле, что эти интегралы равны друг другу) и коммутативность операции свёртка (то, что $(f_1*f_2)(x)=(f_2*f_1)(x)$) фактически уже доказаны при нахождении плотности вероятности случайной величины $Z=X+Y$.

Итак, получено *правило свёртки*: плотность вероятности суммы двух независимых случайных величин равна свёртке их плотностей вероятностей: $f_Z(x)=(f_1*f_2)(x)$.

Симметричное треугольное распределение (Симпсона)

С помощью правила свёртки найдём плотность вероятности суммы двух независимых, равномерно распределённых на промежутке $[a;b)$, случайных величин. Плотности вероятности слагаемых равны: $f_1(x)=f_2(x)=\begin{cases}0,\text{если}\,x<a,\\ \dfrac{1}{b-a},\text{если}\,a<x<b,\\ 0,\text{если}\,x>b.\end{cases}$ Тогда по правилу свёртки (в силу того, что «содержательные части» плотностей вероятностей константы, можно было бы считать площади, что фактически было бы доказательством правила свертки для этого частного случая):

$$f_Z(x)=(f_1*f_2)(x)=\int\limits_{-\infty}^{+\infty}f_1(x-t)f_2(t)dt=$$

$$=\int\limits_{-\infty}^{+\infty}\left(\begin{cases}0,\text{если}\,x-t<a,\\ \dfrac{1}{b-a},\text{если}\,a<x-t<b,\\ 0,\text{если}\,x-t>b.\end{cases}\right)\left(\begin{cases}0,\text{если}\,t<a,\\ \dfrac{1}{b-a},\text{если}\,a<t<b,\\ 0,\text{если}\,t>b.\end{cases}\right)dt=\int\limits_{-\infty}^{a}\left(\begin{cases}0,\text{если}\,t>x-a,\\ \dfrac{1}{b-a},\text{если}\,x-a>t>x-b,\\ 0,\text{если}\,t<x-b.\end{cases}\right)0dt+$$

$$+\int\limits_{a}^{b}\left(\begin{cases}0,\text{если}\,t>x-a,\\ \dfrac{1}{b-a},\text{если}\,x-a>t>x-b,\\ 0,\text{если}\,t<x-b.\end{cases}\right)\frac{1}{b-a}dt+\int\limits_{b}^{+\infty}\left(\begin{cases}0,\text{если}\,t>x-a,\\ \dfrac{1}{b-a},\text{если}\,x-a>t>x-b,\\ 0,\text{если}\,t<x-b.\end{cases}\right)0dt=$$

$$=0+\frac{1}{b-a}\begin{cases}0,\text{если}\,x-a\le a,\\ \displaystyle\int\limits_{a}^{x-a}\frac{dt}{b-a},\text{если}\,a<x-a\le b,\\ \displaystyle\int\limits_{x-b}^{b}\frac{dt}{b-a},\text{если}\,a<x-b<b,\\ 0,\text{если}\,x-b\ge b.\end{cases}+0=\frac{1}{b-a}\begin{cases}0,\text{если}\,x\le 2a,\\ \dfrac{1}{b-a}t\Big|_{a}^{x-a},\text{если}\,2a<x\le a+b,\\ \dfrac{1}{b-a}t\Big|_{x-b}^{b},\text{если}\,a+b<x<2b,\\ 0,\text{если}\,x\ge 2b.\end{cases}=$$

$$=\frac{1}{(b-a)^2}\begin{cases}0,\text{если}\, x\le 2a,\\ x-a-a,\text{если}\, 2a<x\le a+b,\\ b-(x-b),\text{если}\, a+b<x<2b,\\ 0,\text{если}\, x\ge 2b.\end{cases}=\frac{1}{(b-a)^2}\begin{cases}0,\text{если}\, x\le 2a,\\ x-2a,\text{если}\, 2a<x\le a+b,\\ 2b-x,\text{если}\, a+b<x<2b,\\ 0,\text{если}\, x\ge 2b.\end{cases}$$ Чтобы получить функцию распределения такой случайной величины (Симпсона) следует проинтегрировать найденную плотность вероятности:

$$F_Z(x)=\int_{-\infty}^{x}f(t)dt=\frac{1}{(b-a)^2}\begin{cases}\int_{-\infty}^{x}0dt,\text{если}\, x\le 2a,\\ \int_{-\infty}^{2a}0dt+\int_{2a}^{x}(t-2a)dt,\text{если}\, 2a<x\le a+b,\\ \int_{-\infty}^{2a}0dt+\int_{2a}^{a+b}(t-2a)dt+\int_{a+b}^{x}(2b-t)dt,\text{если}\, a+b<x<2b,\\ \int_{-\infty}^{2a}0dt+\int_{2a}^{a+b}(t-2a)dt+\int_{a+b}^{2b}(2b-t)dt+\int_{2b}^{+\infty}0dt,\text{если}\, x\ge 2b.\end{cases}=$$

$$=\frac{1}{(b-a)^2}\begin{cases}0,\text{если}\, x\le 2a,\\ 0+\left.\left(\frac{t^2}{2}-2at\right)\right|_{2a}^{x},\text{если}\, 2a<x\le a+b,\\ 0+\left.\left(\frac{t^2}{2}-2at\right)\right|_{2a}^{a+b}+\left.\left(2bt-\frac{t^2}{2}\right)\right|_{a+b}^{x},\text{если}\, a+b<x<2b,\\ 0+\left.\left(\frac{t^2}{2}-2at\right)\right|_{2a}^{a+b}+\left.\left(2bt-\frac{t^2}{2}\right)\right|_{a+b}^{2b}+0,\text{если}\, x\ge 2b.\end{cases}=$$

$$=\frac{1}{(b-a)^2}\begin{cases}0,\text{если}\, x\le 2a,\\ \left(\frac{x^2}{2}-2ax\right)-\left(\frac{4a^2}{2}-4a^2\right),\text{если}\, 2a<x\le a+b,\\ \left(\frac{(a+b)^2}{2}-2a(a+b)\right)-\left(\frac{4a^2}{2}-4a^2\right)+\\ +\left(2bx-\frac{x^2}{2}\right)-\left(2b(a+b)-\frac{(a+b)^2}{2}\right),\text{если}\, a+b<x<2b,\\ \left(\frac{(a+b)^2}{2}-2a(a+b)\right)-\left(\frac{4a^2}{2}-4a^2\right)+\\ +\left(4b^2-\frac{4b^2}{2}\right)-\left(2b(a+b)-\frac{(a+b)^2}{2}\right),\text{если}\, x\ge 2b.\end{cases}=$$

$$=\frac{1}{(b-a)^2}\begin{cases}0,\text{если}\, x\le 2a,\\ \frac{x^2}{2}-2ax+2a^2,\text{если}\, 2a<x\le a+b,\\ (a+b)^2-4ab-2b^2+2bx-\frac{x^2}{2},\text{если}\, a+b<x<2b,\\ (a+b)^2-4ab,\text{если}\, x\ge 2b.\end{cases}=$$

$$
= \frac{1}{(b-a)^2} \begin{cases} 0, \text{если}\, x \le 2a, \\ \dfrac{x^2}{2} - 2ax + 2a^2, \text{если}\, 2a < x \le a+b, \\ (a-b)^2 - 2b^2 + 2bx - \dfrac{x^2}{2}, \text{если}\, a+b < x < 2b, \\ (a-b)^2, \text{если}\, x \ge 2b. \end{cases} =
$$

$$
= \begin{cases} 0, \text{если}\, x \le 2a, \\ \dfrac{1}{(b-a)^2}\left(\dfrac{x^2}{2} - 2ax + 2a^2\right), \text{если}\, 2a < x \le a+b, \\ 1 - \dfrac{1}{(b-a)^2}\left(\dfrac{x^2}{2} - 2bx + 2b^2\right), \text{если}\, a+b < x < 2b, \\ 1, \text{если}\, x \ge 2b. \end{cases}
$$

На рисунках 37 и 38 показаны графики плотности вероятности и функции распределения симметричного треугольного распределения (Симпсона) при $2a$=0, $2b$=2. Вспомним, что сумма бесконечного числа (в частности, одинаково распределённых) случайных величин, имеет нормальное распределение. Двум слагаемым конечно очень далеко до бесконечной суммы, но посмотрите, насколько график функции распределения для симметричного треугольного распределения похож на график функции распределения для нормального распределения. Причём, обратите внимание, что кривая, полученная стыковкой двух прямых лучей и двух криволинейных отрезков парабол, даже не является ломаной. Все линии в точках стыковок имеют общие касательные: параболы стыкуются с горизонтальными прямыми в своих вершинах, а друг с другом в точке симметрии всей кривой.

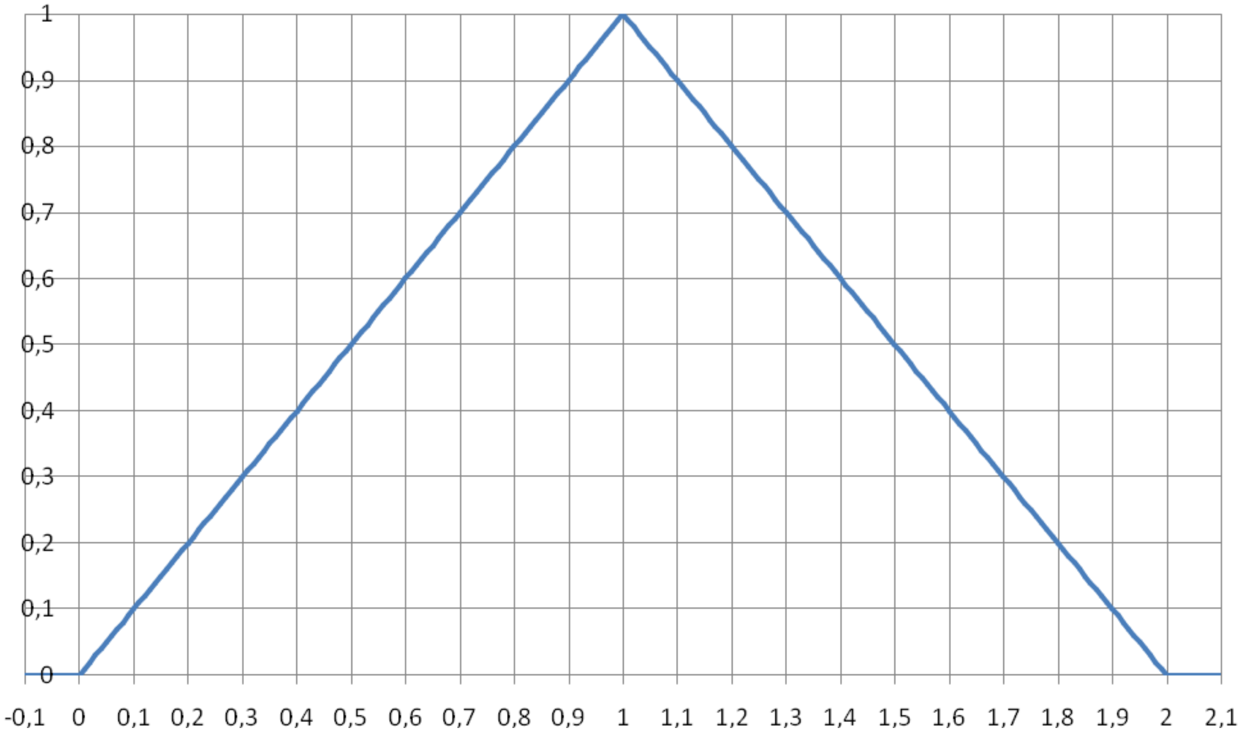

Рис. 37. Плотность вероятности симметричного треугольного распределения (Симпсона) на промежутке [0;2).

Нетрудно найти математическое ожидание и дисперсию распределения Симпсона на промежутке $[2a;2b)$, как суммы математический ожиданий и дисперсий (в силу независимости), соответственно: $EZ=\frac{a+b}{2}+\frac{a+b}{2}=\frac{2a+2b}{2}$, $DZ=\frac{(b-a)^2}{12}+\frac{(b-a)^2}{12}=\frac{(b-a)^2}{6}=\frac{(2b-2a)^2}{24}$, $\sigma(Z)=\sqrt{DZ}=\frac{2b-2a}{2\sqrt{6}}$.

Устойчивость нормального распределения

С помощью формулы для нахождения плотности вероятности суммы двух (возможно зависимых) случайных величин можно найти плотность вероятности суммы нормально распределённых случайных величин X и Y с параметрами $EX=a_1$, $\sigma(X)=\sigma_1$, $EY=a_2$, $\sigma(Y)=\sigma_2$ и корреляцией $\rho(X,Y)=\rho$. Плотность вероятности двумерной случайной величины равна: $f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\cdot\exp\left(\frac{1}{(1-\rho^2)}\cdot\left(-\frac{(x-a_1)^2}{2\sigma_1^2}+\frac{\rho(x-a_1)(y-a_2)}{\sigma_1\sigma_2}-\frac{(y-a_2)^2}{2\sigma_2^2}\right)\right)$. Тогда по общей формуле, справедливой и при зависимости слагаемых, плотность вероятности случайной величины $Z=X+Y$, опуская вычисления, равна:

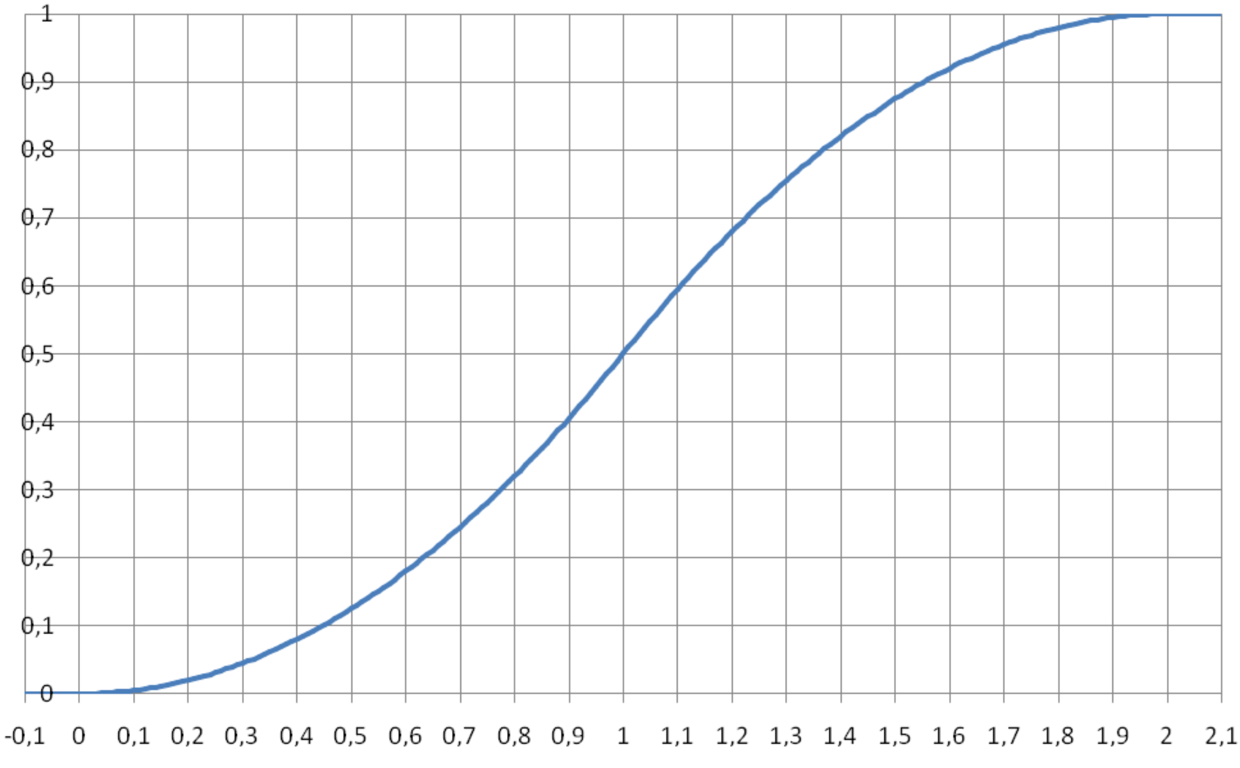

Рис. 38. Функция распределения симметричного треугольного распределения (Симпсона) на промежутке [0;2).

$$f_Z(x)=\int_{-\infty}^{+\infty}f(x-t,t)dt=$$

$$=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\int_{-\infty}^{+\infty}\exp\left(\frac{1}{(1-\rho^2)}\cdot\left(-\frac{(x-t-a_1)^2}{2\sigma_1^2}+\frac{\rho(x-t-a_1)(t-a_2)}{\sigma_1\sigma_2}-\frac{(t-a_2)^2}{2\sigma_2^2}\right)\right)dt=\ldots=$$

$$=\frac{1}{\sqrt{\sigma_1^2+2\rho\sigma_1\sigma_2+\sigma_2^2}}\cdot\frac{1}{\sqrt{2\pi}}\cdot\exp\left(-\frac{(x-(a_1+a_2))^2}{2(\sigma_1^2+2\rho\sigma_1\sigma_2+\sigma_2^2)}\right)=$$

$$= \frac{1}{\sqrt{\sigma_1^2 + 2\rho\sigma_1\sigma_2 + \sigma_2^2}} \cdot \varphi\left(\frac{x-(a_1+a_2)}{\sqrt{\sigma_1^2 + 2\rho\sigma_1\sigma_2 + \sigma_2^2}} \right).$$ Эта функция есть плотность нормального распределения, у которого $EZ=a_1+a_2$, $DZ = \sigma_1^2 + 2\rho\sigma_1\sigma_2 + \sigma_2^2$, $\sigma(Z) = \sqrt{\sigma_1^2 + 2\rho\sigma_1\sigma_2 + \sigma_2^2}$. Если бы случайные величины X и Y были бы независимые, то было бы ρ=0. И формулы $DZ = \sigma_1^2 + \sigma_2^2$ и $\sigma(Z) = \sqrt{\sigma_1^2 + \sigma_2^2}$ соответствовали бы тому, что дисперсия суммы независимых случайных величин равна сумме их дисперсий. А при зависимости случайных величин X и Y получается соответствие формуле $D(X+Y)=DX+2\cdot\mathrm{cov}(X,Y)+DY$.

Компьютерное моделирование случайных величин

В популярную программу Microsoft Office Excel встроена функция СЛЧИС() без аргументов, которая возвращает значение случайной величины равномерно распределённой на промежутке [0;1).

Если необходимо смоделировать другую случайную величину, то следует осуществить пересчёт. Пусть нужно смоделировать результат шахматной партии. Из опыта известно, что около 40% шахматных партий гроссмейстеров заканчиваются победой белых, 25% победой чёрных и 35% партий заканчиваются вничью (партии малоквалифицированных шахматистов редко заканчиваются ничьей из-за большого количества грубых просмотров-зевков). Используем в Microsoft Office Excel функцию: ЕСЛИ(логическое выражение; значение если истина; значение если ложь). У этой функции три аргумента разделённые точкой с запятой (в Microsoft Office Excel разделительным знаком является точка с запятой). Первый аргумент может являться неравенством, которое может удовлетворяться, а может и быть нет, второй аргумент это значение, которое будет возвращено, если это неравенство окажется верным. А третий аргумент то значение, которое будет возвращено, если неравенство окажется невыполненным. В качестве случайной величины возьмём число очков, набранное белыми в имитируемой партии (в шахматах за победу в партии даётся одно очко, за ничью полочка, за поражение нуль очков). Отождествляя экспериментально полученную частоту и вероятность, построим таблицу распределения дискретной случайной величины:

x_i:	0	0,5	1
p_i:	0,4	0,35	0,25

Учитывая, что в Microsoft Office Excel запись формул должна начинаться со знака равенства, в ячейку A1 (без точки, запятой или какого-либо другого знака в конце!) запишем: =СЛЧИС(), а в ячейку B1 формулу для вычисления моделируемой случайной величины: =ЕСЛИ(A1<0,4;1;ЕСЛИ(A1<0,75;0,5;0)). Здесь 0,75 это сумма вероятностей победы белых и ничьи: 0,4+0,35=0,75. Скопировав ячейки A1 и B1 в ячейки, находящиеся ниже, можно получить результаты партий смоделированного шахматного матча.

Пусть необходимо смоделировать непрерывно распределённую случайную величину. В этом случае нужно знать её функцию распределения $F(x)$. По $F(x)$ следует найти обратную ей функцию. Назовём её $G(x)$. Область определения функции $G(x)$ совпадёт со множеством значений функции распределения $F(x)$, то есть будет отрезком [0;1], то есть в качестве аргумента функции $G(x)$ можно использовать встроенную в программу Microsoft Office Excel функцию СЛЧИС(). Результат такой сложной функции и будет моделировать непрерывно распределённую случайную величину с функцией распределения $F(x)$. Докажем это. Обозначим X случайную величину, равномерно распределённую на промежутке [0;1). Тогда её функция распределения $H(x)=\begin{cases}0, \text{если } x<0,\\ x, \text{если } 0<x<1,\\ 1, \text{если } x>1.\end{cases}$ Пусть $Y=G(X)$ однозначно вычисляется при наличии X. Тогда функция распределения наведённой случайной величины Y действительно будет равна: $P(Y<x)=P(G(X)<x)=$ $=P(F(G(X))<F(x))=P(X<F(x))=H(F(x))=F(x)$. В преобразованиях было использовано, что $F(x)$ функция неубывающая (не поменял направление знак неравенства) и то, что композиция исходной и обратной функций есть тождественная функция. Последний переход справедлив потому, что $F(x)\in[0;1]$, а при значении аргумента от нуля до единицы значение функции $H(x)$ совпадает с её аргументом.

Разберем несколько примеров. Если необходимо смоделировать случайную величину равномерно распределённую на промежутке $[a;b)$, то при условии $0<F(x)<1$ верно, что $F(x)=\dfrac{x-a}{b-a}\Leftrightarrow x=a+(b-a)F(x)$. Формально заменяя $F(x)$ на аргумент x получим обратную функцию $G(x)=a+x(b-a)$. Если предположить, что число a хранится в ячейке F1, а число b в ячейке F2, то для получения значения равномерно распределённой на промежутке $[a;b)$ случайной величины в любую свободную ячейку таблицы Excel (например, в C1) следует записать =$F$1+A1*($F$2-$F$1). Знаки доллара нужны для того, чтобы при возможном копировании содержимого ячейки C1 программа Excel воспринимала ссылки на адреса F1 и F2 как абсолютные. Напомним, что в ячейке A1 уже записано: =СЛЧИС().

Если необходимо смоделировать экспоненциальную случайную величину с параметром λ, то при условии $F(x)>0$ верно, что $F(x)=1-e^{-\lambda x}\Leftrightarrow e^{-\lambda x}=1-F(x)\Leftrightarrow -\lambda x=\ln(1-F(x))\Leftrightarrow x=-\dfrac{\ln(1-F(x))}{\lambda}$. Формально заменяя $F(x)$ на аргумент x получим обратную функцию $G(x)=-\dfrac{\ln(1-x)}{\lambda}$. Если предположить, что число λ хранится в ячейке G1, то для получения значения экспоненциальной случайной величины в любую свободную ячейку таблицы Excel (например, в D1) следует записать =-ln(1-A1)/$G$1.

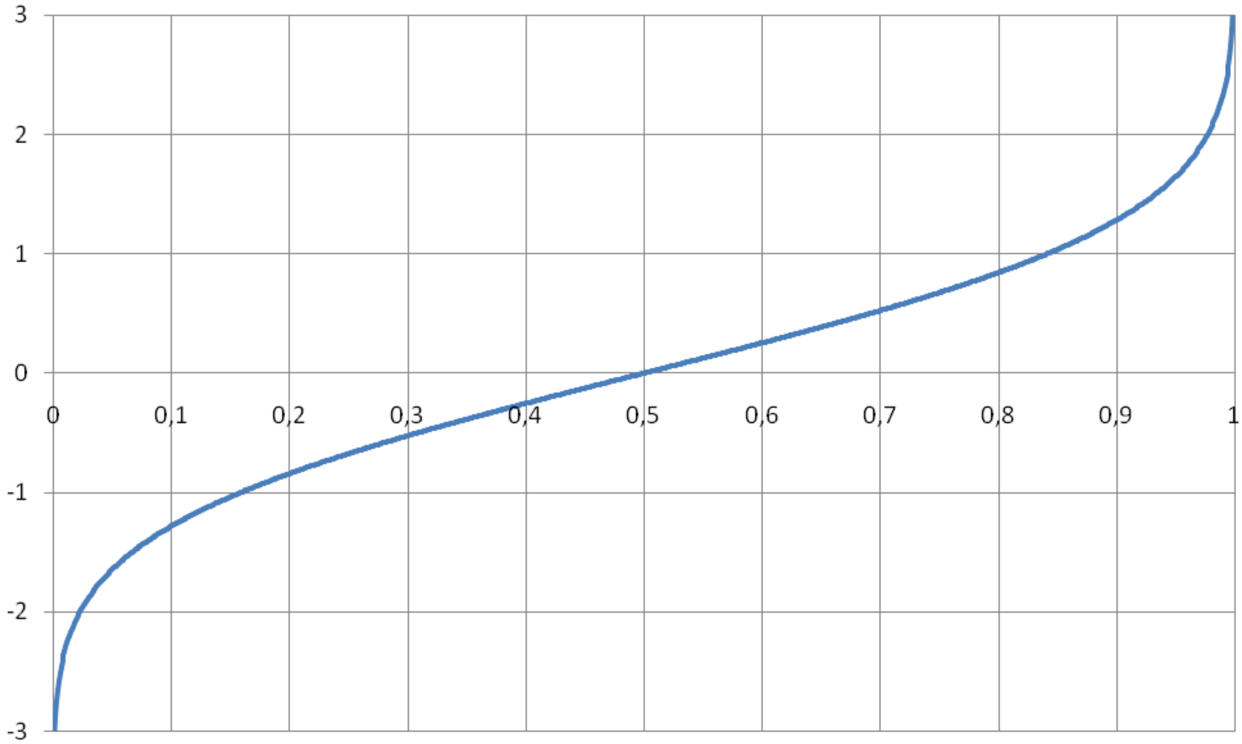

Рис. 39. График функции, обратной к функции распределения стандартной нормальной случайной величины.

Чаще других на практике встречаются случайные величины, распределения которых приближённо можно считать нормальными. Поэтому желательно научиться моделировать нормально распределённую случайную величину. Для этого нужно знать функцию, обратную для функции распределения $F(x) = \Phi\left(\frac{x-a}{\sigma}\right)$ нормально распределённой случайной величины с параметрами a и σ. Сложность заключается в том, что функция $\Phi(x)$ не выражается через элементарные функции и непонятно как вычислять функцию, обратную к ней. К счастью, об этом позаботились разработчики программы Microsoft Office Excel, в которую встроена функция одного аргумента НОРМСТОБР(число), являющаяся обратной для функции $\Phi(x)$. Обозначим эту функцию $\Psi(x)$. Её график изображён на рисунке 39. Заметим, что по одному из свойств взаимообратных функций график функции $\Psi(x)$ может быть получен из графика функции $\Phi(x)$ симметрией относительно биссектрисы первого и третьего координатных углов. Найдём обратную функцию для функции распределения: $F(x) = \Phi\left(\frac{x-a}{\sigma}\right) \Leftrightarrow \Psi(F(x)) = \frac{x-a}{\sigma} \Leftrightarrow x = a + \sigma \cdot \Psi(F(x))$. Формально заменяя $F(x)$ на аргумент x получим обратную функцию $G(x)=a+\sigma\cdot\Psi(x)$. Если предположить, что параметры a и σ хранятся в ячейках H1 и H2 соответственно, то для получения значения нормальной случайной величины в любую свободную ячейку таблицы Excel (например, в E1) следует записать =$H$1+$H$2*НОРМСТОБР(A1).

Если нет возможности использовать Excel со всеми встроенными в него функциями (например, при программировании на каком-нибудь популярном алгоритмическом языке Pascal, Delphi или C++), то возникает достаточно неприятная математическая проблема самостоятельного описания функции, обратной к функции $\Phi(x)$, не являющейся элементарной и задаваемой в виде ряда (пусть даже быстро сходящегося). Эту проблему можно обойти с помощью метода Мюллера. Нужно два раза сгенерировать случайные числа, получив при этом две независимые равномерно распределённые на отрезке [0;1] случайные величины X и Y, а потом вычислить наведённую случайную величину по формуле $Z = \sqrt{-2\ln X}\cdot\cos(\pi Y)$. Заметим, что не следует смущаться знаку «минус» под квадратным корнем, поскольку при X<1 окажется ln$X$<0. Эта случайная величина будет распределена нормально с математическим ожиданием EZ=0 и дисперсией DZ=1. Докажем это. Найдём её функцию распределения.

$$F(x) = P(Z < x) = P\left(\sqrt{-2\ln X}\cdot\cos(\pi Y) < x\right)$$

Рассмотрим первый случай x<0.

Тогда для выполнения неравенства, находящегося под знаком вероятности, обязательно чтобы было cos(πY)<0, что при 0<Y<1 равносильно требованию Y>½. Учитывая, что при делении на отрицательное число знак неравенства переворачивается, получим:

$$F(x) = P\left(\sqrt{-2\ln X} > \frac{x}{\cos(\pi Y)}\right) =$$

поскольку обе части неравенства неотрицательны, то

$$= P\left(-2\ln X > \frac{x^2}{\cos^2(\pi Y)}\right) = P\left(\ln X < -\frac{x^2}{2\cos^2(\pi Y)}\right) = P\left(X < e^{-\frac{x^2}{2\cos^2(\pi Y)}}\right)$$

Изобразим на координатной плоскости ОXY график функции $X = e^{-\frac{x^2}{2\cos^2(\pi Y)}}$ (рис. 40).

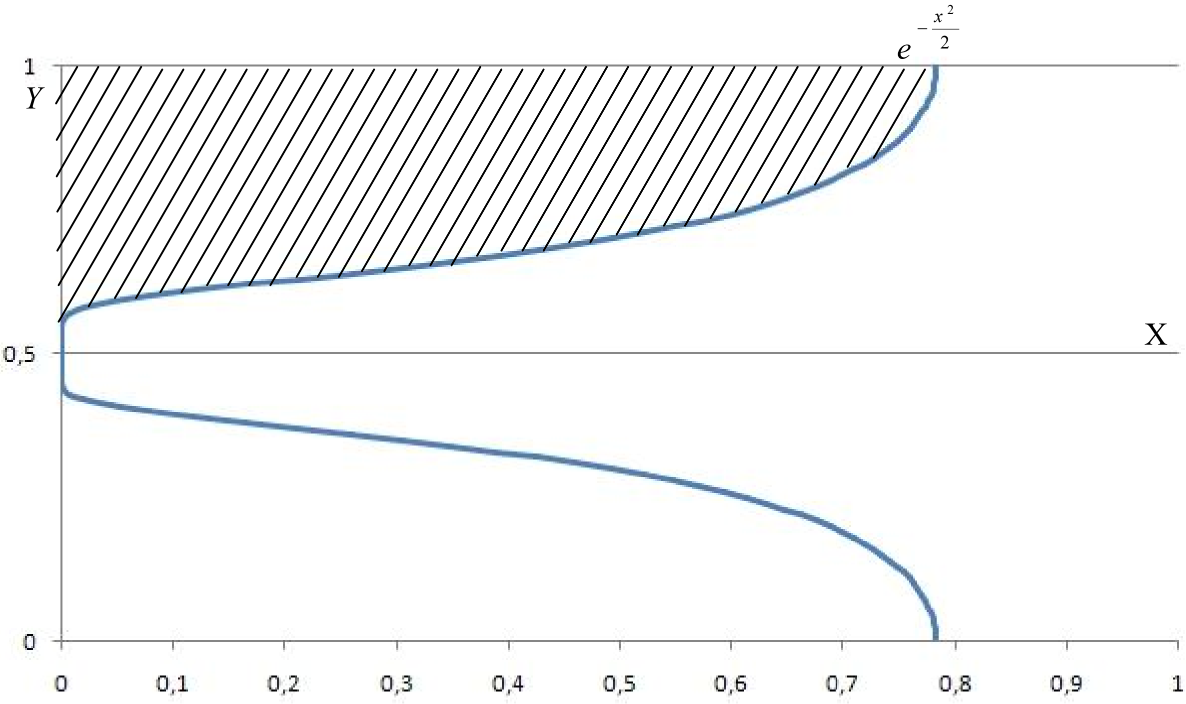

Рис. 40. График функции $X = e^{-\frac{x^2}{2\cos^2(\pi Y)}}$.

Из независимости случайных величин X и Y и того, что эти распределения равномерные, оказываются справедливыми положения геометрической вероятности. Учитывая, что Y>½ искомая вероятность будет равна отношению площади находящейся выше линии Y=½ и левее графика к площади квадрата 1х1 (равной единице). Такую площадь можно выразить определённым интегралом:

$F(x)=\int\limits_{½}^{1} e^{-\frac{x^2}{2\cos^2(\pi Y)}}dY$. Чтобы доказать, что эта функция является функцией распределения нормальной случайной величины с математическим ожиданием 0 и дисперсией 1 можно доказать, что её производная (то есть, плотность вероятности) является функцией $\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$.

$$f(x)=F'(x)=\int\limits_{½}^{1}\left(e^{-\frac{x^2}{2\cos^2(\pi Y)}}\right)'dY=\int\limits_{½}^{1}e^{-\frac{x^2}{2\cos^2(\pi Y)}}\left(-\frac{x}{\cos^2(\pi Y)}\right)dY=-\int\limits_{½}^{1}e^{-\frac{x^2}{2}\left(1+\mathrm{tg}^2(\pi Y)\right)}\cdot\frac{x}{\cos^2(\pi Y)}dY=$$

$$=-\int\limits_{½}^{1}e^{-\frac{x^2}{2}}\cdot e^{-\frac{x^2\mathrm{tg}^2(\pi Y)}{2}}\cdot\frac{x}{\cos^2(\pi Y)}dY=-e^{-\frac{x^2}{2}}\cdot\int\limits_{½}^{1}e^{-\frac{x^2\mathrm{tg}^2(\pi Y)}{2}}\cdot\frac{x}{\cos^2(\pi Y)}dY$$

Сделаем замену переменной в определённом интеграле $t=x\cdot\mathrm{tg}(\pi Y)\Rightarrow dt=\frac{\pi\, x}{\cos^2(\pi Y)}dY$

Пересчитаем пределы интегрирования. При стремлении Y к ½ с правой стороны tg(πY) будет стремиться к минус бесконечности, а учитывая отрицательность x, новая переменная t к плюс бесконечности. При стремлении Y к единице t стремится к нулю.

$$f(x)=-e^{-\frac{x^2}{2}}\cdot\int\limits_{+\infty}^{0}e^{-\frac{t^2}{2}}\cdot\frac{1}{\pi}dt=e^{-\frac{x^2}{2}}\cdot\frac{1}{\pi}\cdot\int\limits_{0}^{+\infty}e^{-\frac{t^2}{2}}dt=$$

поменяв местами пределы интегрирования с изменением знака определённого интеграла и, используя условие нормированности, получим искомый результат:

$$=e^{-\frac{x^2}{2}}\cdot\frac{1}{\pi}\cdot\int\limits_{0}^{+\infty}\frac{\sqrt{2\pi}}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt=e^{-\frac{x^2}{2}}\cdot\frac{\sqrt{2\pi}}{\pi}\cdot\int\limits_{0}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt=e^{-\frac{x^2}{2}}\cdot\frac{\sqrt{2\pi}}{\pi}\cdot\int\limits_{0}^{+\infty}\varphi(t)dt=$$

$$=e^{-\frac{x^2}{2}}\cdot\frac{\sqrt{2\pi}}{\pi}\cdot\frac{1}{2}=e^{-\frac{x^2}{2}}\cdot\frac{1}{\sqrt{2\pi}}=\varphi(x)$$

Рассмотрим второй случай x>0 (в силу парадокса нулевой вероятности случай x=0 можно не рассматривать). Напомним, что $F(x)=P(Z<x)=P\left(\sqrt{-2\ln X}\cdot\cos(\pi Y)<x\right)$

Тогда при cos(πY)<0, что при 0<Y<1 равносильно требованию Y>½ неравенство под знаком вероятности верное, а при cos(πY)>0 то есть при Y<½

$$F(x)=P\left(\sqrt{-2\ln X}<\frac{x}{\cos(\pi Y)}\right)=P\left(-2\ln X<\frac{x^2}{\cos^2(\pi Y)}\right)=P\left(\ln X>-\frac{x^2}{2\cos^2(\pi Y)}\right)=$$

$$=P\left(X>e^{-\frac{x^2}{2\cos^2(\pi Y)}}\right)$$

Эта вероятность будет равна сумме площадей прямоугольника выше линии Y=½ и фигуры справа от графика в зоне, где Y<½. Эта сумма площадей будет равна разности площади всего квадрата 1х1 и части фигуры слева от графика в зоне, где Y<½ (или в силу симметрии графика относительно прямой Y=½ где Y>½, которую считали в первом случае). Поэтому

$$F(x)=1-F(-x)=1-\Phi(-x)=1-\left(1-\Phi(x)\right)=1-1+\Phi(x)=\Phi(x)$$

Что окончательно доказывает справедливость идеи метода Мюллера.

Рекомендуемая литература

1. Агапов Г.И. Задачник по теории вероятностей. М.: Высшая школа, 1994.
2. Бородин А.Н. Элементарный курс теории вероятностей и математической статистики. СПб: Лань, 2006.
3. Вентцель Е.С. Теория вероятностей (первые шаги). М., Физматгиз, 1977.
4. Вентцель Е.С. Теория вероятностей. М.: Наука, 1969.
5. Вентцель Е.С., Овчаров Л.А. Прикладные задачи теории вероятностей. М., Радио и связь, 1983.
6. Гмурман В.Е. Теория вероятностей и математическая статистика. М.: Высшая школа, 2003.
7. Доугерти К. Введение в эконометрику. М.: ИНФРА-М, 2009.
8. Салтыкова Н.М., Сахаров В.Ю. Контрольные задания по теории вероятностей для естественных факультетов. СПб.: изд-во СПбГУ, 1998.
9. Салтыкова Н.М., Сахаров В.Ю. Сборник задач по элементарной теории вероятностей. СПб.: изд-во СПбГУ, 2006.
10. Сборник задач по теории вероятностей, математической статистике и теории случайных функций./ Под ред. А.А.Свешникова. М., Физматгиз, 1965.
11. Секей Г. Парадоксы в теории вероятностей и математической статистике. - М.: Мир, 1990.
12. Феллер В. Введение в теорию вероятностей и её приложения. Т. 1 - 2, - М.: Мир, 1984.
13. Худсон Д. Статистика для физиков. - М.: Мир, 1967.

Содержание